Essays in Financial Econometrics: GMM and Conditional Heteroscedasticity

Mike Aguilar

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2008

Approved by:

Eric Renault, Advisor

Eric Ghysels

Denis Pelletier

Jonathan Hill

William Parke

Abstract

MIKE AGUILAR: Essays in Financial Econometrics: GMM and Conditional Heteroscedasticity.
(Under the direction of Eric Renault.)

This dissertation consists of three papers in the field of financial econometrics. In the first paper, I use a factor structure to model a system of conditionally heteroscedastic asset returns. In the second paper, I illustrate how standard asymptotic results for GMM estimators may be maintained even in the face of moment conditions with infinite variance. In the third paper, I describe a test to distinguish GARCH from Stochastic Volatility models.

Acknowledgments

I would like to thank very much the members of my dissertation committee including Eric Ghysels, Jonathan Hill, William Parke, and Denis Pelletier for useful comments throughout earlier drafts of this dissertation. I would also like to thank Prosper Dovonon for useful conversations regarding Chapter 2 of this dissertation. Conversations with members of the UNC Financial Econometrics Workshop, including Xilong Chen and Jungyeon Yoon, also proved particularly fruitful. Most especially, I would like to thank my advisor Eric Renault, for years of invaluable tutelage and guidance. Of course, all errors herein are my own.

On a personal note, I did not undertake this endeavor alone. I would like to thank my wife for her understanding and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This document presents the three papers that form my dissertation in accordance with the Graduate School and Economics Department at UNC Chapel Hill.

The first paper is titled "Latent Factor Modeling of Multivariate Conditional Heteroscedasticity", wherein I examine the joint dynamics of a system of asset returns. I describe and implement a multivariate factor stochastic volatility (MVFSV) model.

I follow closely the work of Doz and Renault (2006), with two important changes. First, I design a sequential testing procedure to determine the dimensions of the appropriate factor structure needed to accommodate the conditional heteroscedasticity among a system of returns. Second, I employ a form of Tikhonov regularization in order to overcome a near singularity among the moment conditions used for estimation.

Simulation studies suggest that the MVFSV model is able to recover accurately the latent factors that drive the conditional volatility of returns. Moreover, the model estimates can be used to construct conditionally homoscedastic portfolios as linear combinations of the conditionally heteroscedastic assets.

An empirical application to portfolios representing the twelve sectors of the U.S. economy finds that the MVFSV model has important investment implications. Over the period 1993 through 2006, a dynamic asset allocation strategy implied by the MVFSV model is able to perform comparably with a strategy implied by one of the current leaders in multivariate volatility modeling, the Dynamic Conditional Correlation model of Engle (2001). This performance is achieved with minimal distributional assumptions, and does so while maintaining

the parsimony of a factor structure.

The second paper is titled "Indirect Inference for Moment Equations with Infinite Variance", and is co-authored with Eric Renault and Jonathan Hill. Here, we address the issue of moment conditions with infinite variance and the detrimental impact this may have on GMM estimation and inference.

Borrowing from the field of Robust Statistics, we propose truncating the moment conditions used for GMM estimation as a potential remedy for the infinite variance problem. We explore a Winsorizing method of truncation, whereby the extreme realizations of the moment conditions are replaced by some pre-defined threshold. We then use an Indirect Inference approach to relate the truncated series to the parameters of interest through a binding function that we can estimate via simulation. Since the truncated moment conditions have by definition a finite variance, standard GMM asymptotic theory applies. In addition, we develop a method for choosing the optimal truncation threshold by exploiting a tradeoff inherent in the components of the variance matrix of the unknown parameters.

Monte Carlo experiments suggest that our truncation procedure is able to restore standard $\sqrt{T}$ Gaussian asymptotics. Moreover, we are able to validate our choice of truncation threshold and illustrate that the resulting parameter estimates are relatively robust to mis-specification of the simulator used during the Indirect Inference procedure.

The third paper is titled "A Moment-Based Test for GARCH against Stochastic Volatility", and is co-authored with Eric Renault. Here, we develop an original way to nest a GARCH(1,1) model within Stochastic Volatility. This nesting enables us to design a test that distinguishes GARCH from Stochastic Volatility models through a GMM-based inference.

Chapter 2

Latent Factor Modeling of Multivariate Conditional Heteroscedasticity

2.1 Introduction

The literature on modeling univariate volatility processes is well established. GARCH and Stochastic Volatility (SV) models of financial assets have shown to be quite capable in this regard. However, it is clear that the joint estimation of a system of returns is warranted. For instance, Figure 2.7 illustrates the cross-sector correlation coefficients within the U.S. equity market over the last 16 years[1]. The fact that the correlation is non-zero and that it is time varying call for a complete, joint estimation of returns.

Standard multivariate GARCH and SV models have shown some success in accommodating the joint dynamic of such systems. However, they fail to identify the sources of volatility (and co-volatility) movements. Multivariate factor models, such as O-GARCH or its SV counterparts, allow for these sources of volatility to be identified quite easily. The goal of this paper is to implement a multivariate factor stochastic volatility (MVFSV) model.

The class of MVFSV models was actually introduced in an ARCH context by Diebold and Nerlove (1989). In that paper, the authors formulated a multivariate model of returns centered

[1] As described in the Empirical Application section of this paper, I use 12 sectors of the U.S. equity market. For each pair of sectors, I compute the time (t) correlation coefficient over the preceding 3 months. I repeat this process for all pairs of sectors. For each time (t) I then average the correlation coefficients, yielding a time series of cross-sector correlations.

about a single common factor. The factor was designed to follow ARCH dynamics. However, the factor was also deemed to be latent, and hence the SV characterization.

The Diebold and Nerlove (1989) paper was extended in several directions by Harvey, Ruiz and Shephard (1994), King, Sentana and Wadhwani (1994), Jacquier, Polson and Rossi (1995), Shephard (1996), Kim, Shephard and Chib (1998), Pitt and Shephard (1999), Aguilar and West (2000), Fiorentini, Sentana and Shephard (2004), and Doz and Renault (2006). See Asai, McAleer and Yu (2006) for an excellent review.

Latent factor models such as these have several advantages over models that attempt a direct characterization of the variance/covariance matrix of returns. First, the factor representation allows the researcher to capture the time variation in the conditional covariance matrix through the movements of a small number of underlying latent factors. This mitigates the proliferation of parameters often seen in multivariate volatility modeling. Second, by characterizing the common movements in returns and volatilities, these factor models fit nicely into an APT framework. Last, the identification of a limited number of directions of risk has practical advantages to portfolio managers and risk specialists.

The main contribution of this paper is the device of a comprehensive methodology for empirical implementation of the MVFSV model described in Doz and Renault (2006), herein referred to as DR. I illustrate the challenge with estimating this model by GMM and attempt to overcome that difficulty via a form of Tikhonov Regularization. Moreover, I refine a testing strategy associated with the model specification by using a sequential procedure to account for a sequence of nested hypotheses.

The MVFSV model is estimated in two phases. Phase 1 estimates the number of latent factors required to accommodate the conditional heteroscedasticity in a system of asset returns. DR accomplish this through an over-identified restrictions test. I modify this approach slightly by introducing a sequential testing procedure, which accounts for the aforementioned series of nested hypotheses. The sequential procedure is shown to offer a very slight power advantage over the one-step DR approach during hypothesis testing against various alternatives. However, the true virtue of the sequential approach is its ability to effectively re-order the conditionally *hetero*scedastic assets in a way that facilitates the construction of auxiliary portfolios. Akin to

the common features work of Engle and Kozicki (1993), these auxiliary portfolios are formed as linear combinations of the conditionally heteroscedastic asset returns and no longer exhibit that common underlying trait; i.e. they are conditionally *homo*scedastic.

Once the dimensions of the factor structure are revealed in Phase 1, Phase 2 estimates the complete MVFSV model through a well chosen set of moment conditions.

Each phase is estimated via GMM. This has the advantage of overcoming the computational challenge associated with an intractable likelihood function usually encountered when estimating multivariate stochastic volatility models. Moreover, the GMM approach avoids a full parametric specification of the distribution of returns. DR also relaxes certain restrictions, to be made precise in the next section, that are typically placed on the variance/covariance matrix of returns.

Simulation studies indicate that Phase 1 of the MVFSV model is able to identify correctly the number of latent factors in a system of conditionally heteroscedastic asset returns. Moreover, the model identifies linear combinations of the assets that no longer exhibit individual ARCH effects and have constant conditional correlation. In addition, Phase 2 provides parameter estimates sufficient to accurately recover the latent factors that drive the conditional volatility of returns.

An empirical application to portfolios representing the twelve sectors of the U.S. economy finds that the MVFSV model is able to accommodate the individual dynamics of returns quite well throughout the sample period. Joint dynamics are accommodated poorly during the early 1990's, but much more capably during the mid - 1990's through 2006.

The MVFSV model also has important out-of-sample investment implications. A dynamic asset allocation strategy implied by the MVFSV model is able to perform comparably with the Dynamic Conditional Correlation (DCC) model of Engle (2001) for much of the investment period examined. In particular, the MVFSV model is able to outperform the DCC model, along several metrics, during the boom of the late 1990's / early 2000's. Moreover, in many cases, the MVFSV model is able to track more accurately the time variation in the covariance of returns than the DCC model. However, the resulting trading activity can erode portfolio performance when transaction costs are high.

This chapter is organized as follows. Section 2.2 details the model. Section 2.3 outlines the two-phase estimation procedure. Section 2.4 describes the results from a simulation study of Phases 1 & 2. Section 2.5 summarizes the empirical application for 12 sector portfolios of the U.S. equity market.

2.2 Model

I examine the dynamics of n asset returns (y_{t+1}) and their (n x n) conditional variance matrix $\Sigma_t = V(y_{t+1}|J_t)$. J_t is an information set that contains past values of the returns, $y_\tau \forall\ \tau \leq t$, as well as the past of some unobserved common factors, $f_\tau \forall\ \tau \leq t$. Notice that J_t differs from the econometrician's information set I_t, which may contain only past values of returns. Specifically, $I_t \subset J_t$. For notational convenience I will denote all relevant information sets through a simple time subscript, so that $\Sigma_t \equiv V_t(y_{t+1}) \equiv V(y_{t+1}|J_t)$.

Consider the decomposition of this variance matrix:

$$\Sigma_t = \Lambda D_t \Lambda' + \Omega_t \tag{2.1}$$

where D_t is a diagonal matrix of size K with diagonal coefficients $\sigma^2_{kt}, k = 1, ..., K$. Consider, for now, Ω_t to be a diagonal positive definite matrix. Moreover, the $\sigma^2_{kt}, k = 1, ..., K$ are positive, stationary stochastic processes with unit expectation and non-zero variance. The unconditional expectation of D_t is an Identity matrix of size K.

Viewing the σ^2_{kt} as conditional variances of K independent common factors, $\sigma^2_{kt} = V_t(f_{k,t+1})$, enables the formation of the K-factor conditional regression representation:

$$y_{t+1} = \mu + \Lambda f_{t+1} + u_{t+1} \tag{2.2}$$

where f_{t+1} is a (K x 1) vector of unobserved latent factors, u_{t+1} is an (n x 1) vector of idiosyncratic terms, Λ is an (n x K) matrix of factor loadings, and μ is an (n x 1) vector of constants that may be interpreted as risk premia if y_{t+1} is taken to be in excess of the risk free rate. Ascribing dynamics to f_{t+1} is unnecessary at this point, but will be discussed in section

2.3.3[2]

I define the variance/covariance matrices of the errors and factors in equation (2.2) as $E_t(u_{t+1}u'_{t+1}) = \Omega_t$ and $E_t(f_{t+1}f'_{t+1}) = D_t$, respectively. I also impose the following assumptions: $E_t(f_{t+1}) = 0$, $E_t(u_{t+1}) = 0$, and $E_t(f_{t+1}u'_{t+1}) = 0$. Notice that this last assumption allows us to interpret the factor loadings as standard conditional regression coefficients of returns on the factors, and $\Omega_t = Var_t[u_{t+1}]$ as the residual risk.

A factor-analytic structure such as that presented above is not new. A majority of the MVFSV models mentioned earlier use this formulation. However, DR part from the previous literature in two ways. First, the matrix of residual risk is deemed to be time invariant. This assumption implies that the conditional heteroscedasticity of returns can be captured completely through the movements of the underlying common factors. Second, the diagonality of the residual risk matrix is not preserved by portfolio formation. As such, DR allow for non-zero off-diagonal elements. The variance/covariance matrix of the returns can then be written as

$$\Sigma_t = \Lambda D_t \Lambda' + \Omega. \tag{2.3}$$

where Ω is a possibly non-diagonal, time invariant matrix.

The price to pay for this specification of Ω is that we are unable to identify uniquely all the parameters in the model. Importantly, only the range of Λ is identified. This is due to the non-diagonal structure permitting any constant part of the conditional variance of the factors to be transferred to the variance of the errors by a simple re-scaling of the loading vectors.

Consider $k \leq K$ as a particular number of latent factors being considered for a system of asset returns. Denote $\overline{y}_{t+1}$ as the first k components of y_{t+1} and $\overline{\overline{y}}_{t+1}$ as the remaining $(n-k)$ elements of y_{t+1}. Consistent with the common features work of Engle and Kozicki (1993), DR show that there exist linear combinations of the asset returns that are conditionally homoscedastic. That is, if there exists a common factor that explains the conditional

[2]Note: DR detail the possibility of a time varying risk premium with the specification $\mu_t = \mu + \Lambda\phi_t$. I leave this for future work.

heteroscedasticity among the assets in this system, there should be linear combinations of those assets that no longer contain conditional heteroscedasticity. Specifically, for a suitable partition of $y = [\overline{y}' \ \overline{\overline{y}}']$, the conditionally homoscedastic portfolios take the form $(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})$, where B is a constant matrix of dimension $((n-k)$ x $k)$.

2.3 Estimation

I estimate the MVFSV model in two phases, each employing a GMM technique. Phase 1 determines the number of latent factors needed to accommodate adequately the temporal dynamics of the assets under consideration. Phase 2 estimates the complete multivariate factor stochastic volatility model.

2.3.1 Phase 1: Search for the Number of Common Factors

The empirical goal of Phase 1 is to determine the number of factors (K). This is a non-trivial task that has been dealt with in a rather ad-hoc fashion in the literature. Broadly speaking, there are three techniques that are used to determine the number of factors in a system of asset returns: 1) Common Features, 2) Principal Components, and 3) Model Selection Criteria. This certainly is not an exhaustive list of options, nor are they necessarily mutually exclusive. However, I use these three to characterize a general class of possibilities for my estimation procedure.

Common Features, as espoused by Engle and Kozicki (1993), appears to be the least utilized of the three major techniques in the field of volatility modeling. Lanne and Saikkonen (2007) offer a rare example of this approach, wherein the authors deal explicitly with the issue of determining the number of latent factors. The authors choose a number of factors a priori and then estimate a factor GARCH model. They then validate this choice post-estimation through an original testing procedure for common features.

Principal Components, on the other hand, is perhaps the most widely used technique to determine the number of factors in the GARCH literature. The O-GARCH model of Alexander (2001) determines the number of factors to be considered pre-estimation. The asset returns

are decomposed into the their (k) principal components, accounting for some pre-determined amount total variation among the returns, say 90%. The author then estimates the GARCH model upon the "factors".

Model Selection Criteria are commonly used in determining the dimension of the factor space in multivariate stochastic volatility models. For instance, Connor, Korajczyk and Linton (2006) choose a range of potential dimensions for the factor space a priori and then estimate their stochastic volatility model over each choice of (k). The appropriate number of factors is then chosen post-estimation. Broadly speaking, I categorize this type of analysis as Model Selection because the authors choose the number of factors that best fits their model to the data, with a penalization term for the number of parameters.

I address the issue of determining the size of the factor space explicitly. Moreover, (k) is selected pre-estimation. I say "pre"-estimation because the dimension of the factor space is determined before the full model is estimated in Phase 2.

DR show that efficient estimation of the stochastic volatility factor structure, and thus the dimension of the factor space, should be conducted via:

$$E_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})y'_{t+1}] = C \tag{2.4}$$

The common features work of Engle and Kozicki (1993) is the obvious motivating force for these moment conditions since they are based upon the conditionally homoscedastic portfolios $(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})$[3]. However, the information content in the stochastic volatility factor structure is greater than that of a pure common features model. For instance, if I partition C in an

[3]The homoscedasticity of this linear combination of assets is seen easily by partitioning the loading vector $\Lambda = [\overline{\Lambda}\ \overline{\overline{\Lambda}}]'$ and rewriting equation (2.2) as

$$\overline{y}_{t+1} = \overline{\Lambda} f_{t+1} + \overline{u}_{t+1} \tag{2.5}$$

$$\overline{\overline{y}}_{t+1} = \overline{\overline{\Lambda}} f_{t+1} + \overline{\overline{u}}_{t+1} \tag{2.6}$$

(Notice that I drop the mean return vector μ for ease of exposition.) When we have k factors, a suitable partition ensures that $\overline{\Lambda}$ is a $(k \times k)$ non-singular matrix. Solving equation (2.5) is then possible, and plugging into equation (2.6) yields:

$$\overline{\overline{y}}_{t+1} - \overline{\overline{\Lambda}}\overline{\Lambda}^{-1}\overline{y}_{t+1} = \overline{\overline{u}}_{t+1} - \overline{\overline{\Lambda}}\overline{\Lambda}^{-1}\overline{u}_{t+1} \tag{2.7}$$

Defining $B = \overline{\overline{\Lambda}}\overline{\Lambda}^{-1}$ and noticing that the left hand side of equation (2.7) is a linear function of the homoscedastic vector u_{t+1} implies that $(\overline{\overline{y}}_{t+1} - B\overline{y})$ is itself conditionally homoscedastic.

obvious fashion, (2.4) can be re-written as

$$E_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})\overline{y}'_{t+1}] = C_1 \tag{2.8}$$

$$E_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})\overline{\overline{y}}'_{t+1}] = C_2 \tag{2.9}$$

Meanwhile, the common features model contains information only about linear combinations of the equations above:

$$E_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})'] = C_2 - BC_1 \tag{2.10}$$

For a convenient choice of H instruments $z_t \in I_t$, the moment conditions

$$E[z_t \otimes vec[E_t(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})y'_{t+1} - C]] = 0 \tag{2.11}$$

pave the way for GMM estimation and inference. The null hypothesis associated with an over-identified restrictions test is that k factors are sufficient to accommodate the conditional heteroscedasticity among the n assets that comprise y_{t+1}. An equivalent phrasing of the null is that there exist $(n-k)$ conditionally homoscedastic portfolios of the form $(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})$.

Upon careful inspection it becomes clear that this null nests a series of hypotheses. Accepting the null of k factors being sufficient to accommodate the conditional heteroscedasticity among the n assets implies that at most k factors are needed to accommodate the dynamics of any subset of the n assets. Failing to test the dimension of the factor structure for subsets of the asset space creates a possible loss of power.[4] To account for these nested hypotheses, I part from DR slightly by incorporating a sequential testing procedure.

Begin by ordering the assets according to the prominence of conditional heteroscedasticity evident in each process. $y^{(1)}$ is the most conditionally heteroscedastic, and $y^{(2)}$ is the least. I then allocate subsets of the asset space to $\overline{y}$ and $\overline{\overline{y}}$ according to the number of factors being considered.

[4]Simulation studies suggest a power loss of 3-5 percentage points from ignoring the sequential procedure.

Consider the possibility of a single common factor. The first step of the sequential procedure asks whether this factor is sufficient to accommodate the conditional heteroscedasticity among the first two assets. I will detail the test statistic shortly. Allocate the first asset, $y^{(1)}$, to $\overline{y}$ and the second asset, $y^{(2)}$, to $\overline{\overline{y}}$. Failing to reject the null hypothesis in the first step of the sequence suggests that a single common factor is capable of accommodating the conditional heteroscedasticity among the first two assets.

The second step of the sequence asks whether this single common factor is also sufficient for the dynamics of the first three assets. In this step, $\overline{y}$ continues to consist of $y^{(1)}$, and $\overline{\overline{y}}$ expands to include both $y^{(2)}$ and $y^{(3)}$. Failing to reject the null is this second step of the sequence suggests that a single common factor is capable of accommodating the conditional heteroscedasticity among the first three assets.

A sample of the sequential testing procedure for a single factor is as follows:

H_o^1 : 1 Factor is sufficient for assets $(y^{(1)}, y^{(2)})$

H_a^1 : 1 Factor is not sufficient for assets $(y^{(1)}, y^{(2)})$

If fail to reject, proceed to next step in sequence.

If reject, stop this sequence and consider two factors.

H_o^2 : 1 Factor is sufficient for assets $(y^{(1)}, y^{(2)}, y^{(3)})$

H_a^2 : 1 Factor is not sufficient for assets $(y^{(1)}, y^{(2)}, y^{(3)})$

If fail to reject, proceed to next step in sequence.

If reject, stop this sequence and consider two factors.

$\vdots$

H_o^{n-1} : 1 Factor is sufficient for assets $(y^{(1)}, y^{(2)}, \ldots, y^{(n)})$

H_a^{n-1} : 1 Factor is not sufficient for assets $(y^{(1)}, y^{(2)}, \ldots, y^{(n)})$

If fail to reject, stop.

If reject, stop this sequence and consider two factors.

Failing to reject the last null in the sequence above suggests that a single common factor is sufficient to accommodate the conditional heteroscedasticity among all n assets. As a result, there exists $n-1$ conditionally homoscedastic portfolios formed as linear combinations of the

n conditionally heteroscedastic primitive assets.

Rejecting the null hypothesis at any stage of the sequence above suggests the possibility of two or more common factors. For instance, assume H_0^2 is rejected. This implies that $y^{(3)}$ introduced new dynamics to the system that could not be accommodated by the single factor that was sufficient for $y^{(1)}$ and $y^{(2)}$.

To account for the additional dynamics, consider a new sequence of tests. The first null hypothesis of this new sequence asks whether two common factors are sufficient for $y^{(1)}$, $y^{(2)}$, and $y^{(3)}$. Once again, allocate the first asset, $y^{(1)}$, to $\overline{y}$. Additionally, expand $\overline{y}$ to include $y^{(3)}$, which is the asset that introduced the new dynamics. The set $\overline{\overline{y}}$ now consists only of $y^{(2)}$. The sequence proceeds in a similar fashion as that outlined above. Failing to reject the first null suggests that two factors are accommodative. Proceed by expanding $\overline{\overline{y}}$ to include more assets. On the other hand, rejecting the first null suggests that two factors are not sufficient and $\overline{y}$ must expand in order to consider three factors.

In this sense, the sequential procedure re-orders the assets in a way that identifies suitable candidates for the $\overline{y}$ vector.

Notice that moving through the sequence of hypotheses is done in a lexicographical manner, allocating, whenever possible, the most conditionally heteroscedastic assets to $\bar{y}$. This is useful for two reasons. First, it limits greatly the possible number of orderings of the assets we need to consider. Second, it allows $\bar{y}$ to capture those assets for which the factors are designed; i.e. those assets with the largest and most unique forms of conditional heteroscedasticity.

As is well documented in the literature, care must be taken when evaluating a sequence of nested hypotheses. Particular attention must be given to the construction of the test statistics, the degrees of freedom, and the size of the test.

The test statistic at any stage of the sequence is the J-statistic from an over-identified restrictions test. Consider the typical GMM objective function $J = Tg'Wg$, where g is the sample mean of the moments, the weighting matrix (W) is set optimally to the inverse of the variance matrix of the moments (S), and T is the sample size. The degrees of freedom is equal to $p - q$, where p is the number of moment conditions and q is the number of parameters in the model. This is all standard. Unfortunately, naively using this test statistic and degrees

of freedom at any given step of the sequence ignores the fact that I have accepted each of the preceding steps. The J-statistic and the degrees of freedom must be adjusted.

The motivation for the adjustment comes from Eichenbaum, Hansen and Singleton (1998). Using their framework, I view each step of the sequence as a test of a subset of the moment conditions. To illustrate in a general setting, consider an H-dimensional instrument vector z. Moreover consider a set of moment conditions that can be partitioned into two non-overlapping subsets a and b as follows:

$$m^a(\theta) = E\left[\begin{array}{c} z \otimes \left[\begin{array}{c} (y^{(2)} - B1y^{(1)})(y^{(1)})' - C1 \\ (y^{(2)} - B1y^{(1)})(y^{(2)})' - C2 \end{array}\right] \end{array}\right] \tag{2.12}$$

where q_a is the number of parameters and p_a is the number of moment conditions;

$$m^b(\theta) = E\left[\begin{array}{c} z \otimes \left[\begin{array}{c} (y^{(2)} - B1y^{(1)})(y^{(3)})' - C3 \\ (y^{(3)} - B2y^{(1)})(y^{(1)})' - C4 \\ (y^{(3)} - B2y^{(1)})(y^{(2)})' - C5 \\ (y^{(3)} - B2y^{(1)})(y^{(3)})' - C6 \end{array}\right] \end{array}\right] \tag{2.13}$$

where q_b is the number of parameters and p_b is the number of moment conditions. Let θ be a parameter vector containing the relevant elements of B & C.

Now consider the hypothesis where subsets a and b hold under the null, but only subset a holds under the alternative.

$$\begin{aligned} H_0^E &: E[m^a(\theta)] = 0 \quad \& \quad E[m^b(\theta)] = 0 \\ H_A^E &: E[m^a(\theta)] = 0 \quad \& \quad E[m^b(\theta)] \neq 0 \end{aligned} \tag{2.14}$$

Eichenbaum, Hansen and Singleton (1998) offer a simple test statistic, which is the difference of the J-statistics from the over-identified restrictions test. Denote $J_a^E = Tg^a(\hat{\theta})'Wg^a(\hat{\theta})$ as the test statistic pertaining only to the subset a of moment conditions, and that which is true under both the null and the alternative. Moreover, denote $J_{a+b}^E = Tg^{a+b}(\hat{\theta})'Wg^{a+b}(\hat{\theta})$ as the test statistic using the entire set of moment conditions. The test statistic accompanying the

joint hypothesis in (2.14) is then $\xi^E = J^E_{a+b} - J^E_a$. Under the proper regularity conditions, this test statistic is chi-square distributed with $q_b - p_b$ degrees of freedom.

To see how this framework can be applied to Phase 1 of the MVFSV model, consider the first two steps of a sequence in a simple three asset case. I will drop time subscripts for notational convenience. The first step can be written as:

$$H^1_0 : 1 \text{ Factor is sufficient for assets } (y^{(1)}, y^{(2)})$$
$$H^1_a : 1 \text{ Factor is not sufficient for assets } (y^{(1)}, y^{(2)})$$

$$m^1(\theta) = E\left[z \otimes \begin{bmatrix} (y^{(2)} - B1y^{(1)})(y^{(1)})' - C1 \\ (y^{(2)} - B1y^{(1)})(y^{(2)})' - C2 \end{bmatrix} \right] \tag{2.15}$$

The second step can be rewritten as

$$H^2_0 : 1 \text{ Factor is sufficient for assets } (y^{(1)}, y^{(2)}, y^{(3)})$$
$$H^2_a : 1 \text{ Factor is not sufficient for assets } (y^{(1)}, y^{(2)}, y^{(3)})$$

$$m^2(\theta) = E\left[z \otimes \begin{bmatrix} (y^{(2)} - B1y^{(1)})(y^{(1)})' - C1 \\ (y^{(2)} - B1y^{(1)})(y^{(2)})' - C2 \\ (y^{(2)} - B1y^{(1)})(y^{(3)})' - C3 \\ (y^{(3)} - B2y^{(1)})(y^{(1)})' - C4 \\ (y^{(3)} - B2y^{(1)})(y^{(2)})' - C5 \\ (y^{(3)} - B2y^{(1)})(y^{(3)})' - C6 \end{bmatrix} \right] \tag{2.16}$$

Moment conditions $m^1(\theta)$ are analogous to $m^a(\theta)$ in the general example, and $m^2(\theta)$ are analogous to $m^{a+b}(\theta)$. Therefore, in order to construct a proper test statistic, I need only subtract the J-statistics and degrees of freedom along each step of the sequence. For example, the adjusted test statistic for step 2 of the sequence is $\xi^2 = J^2 - J^1$. Moreover, the adjusted degrees of freedom is merely the difference in the degrees of freedom from each step along the sequence. Specifically, $J^1 \sim \chi^2_{q_1 - p_1}$, $J^2 \sim \chi^2_{q_2 - p_2}$, and the adjusted test statistic in the second step is $\xi^2 = J^2 - J^1 \sim \chi^2_{q_2 - p_2 - (q_1 - p_1)}$.

Care must also be taken when denoting the nominal size of each step along the sequence. As Gourieroux and Monfort (1995) illustrate nicely, the nominal size of the test compounds as I progress along the sequence, thereby reducing the confidence I can draw from each inference.

The sequence of tests is designed as a descending step-wise procedure. The null in step one is rather broad. The single factor identified need account only for the conditional heteroscedasticity in two assets. If I fail to reject this null, I proceed to the next step of the sequence by adding on the additional requirement that the factor be able to accommodate the conditional heteroscedasticity among three assets. Through the descending step-wise approach, I am able to control for the nominal size of the test along every step of the sequence.

Denote α_j as the level of the test at step j of the sequence. The probability of a Type I error at step j can be seen as a function of the levels for all the preceding steps of the sequence. Specifically, Pr[reject $H_0^j | H_0^j$ true]= 1-$\prod_{i=1}^{j}(1-\alpha_i)$. Setting $\alpha_i = \alpha \; \forall i$, I can induce the significance level for testing H_0^j as $1-(1-\alpha)^j$.

Let us revisit the small-scale example of three assets. The last null in the sequence, and ultimately the one of interest, is H_0^2, which I want to test at the 10% level. Solving for α; $1-(1-\alpha)^2 = 0.10 \rightarrow \alpha = 0.0513$. I can now induce the levels of significance at each step of the sequence: H_0^1 is tested at 0.0513, and H_0^2 is tested at 0.10, as desired.

With the sequential testing procedure in hand, I should be able to identify K, the number of factors in the system of assets. Unfortunately, a difficulty arises when constructing the variance/covariance matrix of the moment conditions, which is required for efficient GMM estimation. By the design of the moment conditions, elements of $\bar{y}$ interact repeatedly with the elements of $\bar{\bar{y}}$, forming linear combinations of the assets that are similar to one another. As a result, the columns of the variance/covariance matrix are nearly redundant, which yields a near-singularity and makes inverting this matrix difficult. I attempt to overcome this difficulty through a form of Tikhonov regularization.

2.3.2 Tikhonov Regularization

I attempt to overcome the aforementioned estimation difficulty through a form Tikhonov regularization. The technique introduces small perturbations to the diagonal elements of

the variance matrix S. These perturbations, or Tikhonov Factors, must be large enough to alleviate the ill-posed problem, yet the smallest possible in order to consider that we still have approximately reach efficient GMM. I regularize S as follows:

$$S^* = \frac{1}{T}\sum_{t=1}^{T} m_t m_t' + \alpha^* I_{j(n-K)n} \tag{2.17}$$

where α^* is the Tikhonov Factor[5]. I then use the following GMM objective function for estimation: $J^* = Tg'W^*g$, where $W^* = (S^*)^{-1}$.

The challenge in utilizing regularization techniques is calibrating the size of the Tikhonov Factor[6]. Picking α^* too small will not avoid the near singularity. Picking α^* too large will cause S^* to grow large, sending the GMM objective function (J) toward zero. This contaminates the parameter estimates and precludes reliable inference[7]. Mindful of this tradeoff, I offer here a somewhat crude, but effective means of choosing α^*.

Begin by choosing a large α^* that allows us to avoid the near singularity issue. Conduct GMM estimation with the weighting matrix equal to $(S^*)^{-1}$. Gather the parameter estimates into vector $\Theta(1)$. Pick another α^* by decreasing the previous choice. Re-estimate and gather the parameter vector $\Theta(2)$. Compute the element-wise percentage change in the parameter vectors and take its norm; $d(2) = \| \frac{\Theta(2)}{\Theta(1)} - 1 \|$. This $d(2)$ is a single point in Figure 2.7. Pick another α^* by decrementing the previous choice. Re-estimate and gather the parameter vector $\Theta(3)$. Compute the element-wise percentage change of the parameter vectors and take its norm; $d(3) = \| \frac{\Theta(3)}{\Theta(2)} - 1 \|$. Repeat this process until $d(i+1)$ rises above a given tolerance; $d(i+1) \equiv \| \frac{\Theta(i+1)}{\Theta(i)} - 1 \| > tol$. The α^* associated with the i^{th} step of this procedure is the appropriate choice for the Tikhonov Factor since it is the smallest one to provide a reliable estimator.

---

[5] I can also form a "regularized" Newey-West variance / covariance matrix as follows: $S^*_{NW} = \widehat{\Gamma}(0) + \sum_{j=1}^{T-1} \kappa(j,b)[\widehat{\Gamma}(j) + \widehat{\Gamma}(j)'] + \alpha^* I_{q(n-K)n}$, where κ is the kernel and b is the bandwidth.

[6] See Carrasco (2007) for an excellent discussion of regularization methods in an instrumental variables context.

[7] In fact, we want inference to be reliable when it is based on standard formulas for efficient GMM as if $\alpha = 0$.

2.3.3 Phase 2: Full Model Estimation

Once the number of factors is identified in Phase 1, I can estimate the complete K-factors conditional regression representation defined in equation (2.2).

Estimation requires a specification for the dynamics of the factor volatilities, for which I assume an affine mean reverting structure:

$$E_{t-1}(\sigma_{kt}^2) = 1 - \gamma_k + \gamma_k \sigma_{kt-1}^2 \tag{2.18}$$

Unfortunately, the latent nature of the factors and the non-diagonal residual risk (Ω) prevents unique identification of the loading vector Λ. In fact, only the range of Λ can be identified (see Fiorentini and Sentana (2001)). This lack of identification in Λ carries over to the other parameters of interest in the model. However, DR show that the stochastic volatility factor structure permits estimation of functions of the parameters that are invariant to scale changes in Λ.

Under the constant risk premium specification, the parameters $B, \gamma_k, k = 1, ..., K$, and $(\Omega_2 - B\Omega_1)$ can be uniquely characterized by the following set of conditional moment restrictions:

$$\begin{aligned} &vecE_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})y'_{t+1}] = vec(\overline{\overline{\widehat{\mu}}} - B\overline{\widehat{\mu}})\widehat{\mu}' + (\Omega_2 - B\Omega_1) \\ &diagE_{t-K}[\textstyle\prod_{k=1}^{K}(1-\gamma_k L)(y_{t+1}y'_{t+1})] = diag(\widehat{\mu}\widehat{\mu}' + \widehat{\Lambda\Lambda' + \Omega})\textstyle\prod_{k=1}^{K}(1-\gamma_k) \end{aligned} \tag{2.19}$$

once a proper set of instruments is chosen. The estimates of the unconditional mean and variance of returns, $\widehat{\mu}$, and $\widehat{\Lambda'\Lambda + \Omega}$, are obtained by their sample averages; $\widehat{\mu} = 1/T \sum_{t=1}^{T} y_t$ and $\widehat{\Lambda'\Lambda + \Omega} = 1/T \sum_{t=1}^{T} (y_t - \widehat{\mu})^2$.

I refer the reader to DR for proofs illustrating the ability of these moment conditions to identify the stated parameters. However, the intuition is rather straightforward. The first set of moment conditions are akin to those used in Phase 1, estimating B and $\Omega_2 - B\Omega_1$. The second set of moment conditions identify γ, and are in the spirit of the multi-period moments conditions of Meddahi and Renault (2004), where the $(1 - \gamma L)$ filter annihilates the dynamics

of σ_t^2.

The moment conditions in equation (2.19) differ somewhat from the DR moment conditions for Phase 2. First, the unconditional mean and variance are estimated via their sample counterparts outside of Phase 2. This reduces the dimension of the parameter space. Similarly, only the diagonal elements of the second moment condition in equation (2.19) are used. DR use the entire lower triangular portion of the moment conditions, which unnecessarily overidentifies the model. In effect, that approach uses covariance terms in the estimation of the variance. Dovonon (2006) also utilizes the diagonal structure of this moment condition.

2.4 Simulation Study

In this section I conduct several simulation exercises to evaluate the efficacy of the MVFSV model. Specifically, I aim to present evidence of the presence of univariate ARCH effects as well as multivariate dynamic condition correlation (DCC) effects in the simulated returns. I then estimate Phase 1 of the estimation strategy using the regularization method. With the estimates of B in hand, I form auxiliary portfolios and test for the presence of conditional homoscedasticity, in both a univariate and multivariate sense. In addition, I use estimates from Phase 2 to extract the latent factors and compare these to the actual simulated factors.

The test I use for examining univariate conditional heteroscedasticity is the standard ARCH LM test due to Engle (1982). Akin to the Breusch-Godfrey test for autocorrelation, the ARCH LM test regresses squared residuals upon their own past. The test statistic is TR^2, and is distributed χ_q^2 where q is the order of the ARCH process.

The test for DCC effects is due to Engle and Sheppard (2001). Consider the $(n \times 1)$ vector of returns $y_t|J_{t-1} \sim N(0, H_t)$, where $H_t \equiv D_t R_t D_t$. D_t is an $(n \times n)$ diagonal matrix of time varying standard deviations from univariate GARCH models. R_t is the potentially time varying correlation matrix. The test of constant conditional correlation is as follows:

$$H_o : R_t = \overline{R} \; \forall \; t \; \epsilon \; T$$
$$H_a : vech(R_t) = vech(R_t) + \beta_1 vech(R_{t-1}) + ... + \beta_p vech(R_{t-p})$$

Test test centers around an accompanying artificial vector auto-regression:

$$Y_t = \alpha + \beta_1 Y_{t-1} + ... + \beta_L Y_{t-L} + \eta_t$$

where $Y_t = vech^u[(\overline{R}^{-1/2} D_t^{-1} y_t)(\overline{R}^{-1/2} D_t^{-1} y_t)' - I_g]$, with $(\overline{R}^{-1/2} D_t^{-1} y_t)$ representing a $(g \times 1)$ vector of returns jointly standardized under the null and $vech^u$ is an operator that selects the elements above the diagonal of a matrix.

Under the null of a constant conditional correlation, the constant and all the lagged parameters in the auxiliary regression should be zero. The accompanying test is conducted via a seemingly unrelated regression with a resulting test statistic of $\frac{\widehat{\delta} X X' \widehat{\delta}'}{\widehat{\sigma^2}}$, which is asymptotically distributed χ^2_{L+1}. The $\widehat{\delta}$ are the estimated regression parameters, X contains all the regressors including the constant, and $\widehat{\sigma^2}$ is the estimated variance of the error term.

2.4.1 Simulating Return Paths

All of the simulation exercises I will undertake are based on equation (2.2), which I re-write here for convenience:

$$y_{t+1} = \mu + \Lambda f_{t+1} + u_{t+1} \tag{2.20}$$

Define y_{t+1} as an $(n \times 1)$ vector of asset returns. Unless otherwise stated, the risk premium, μ, is assumed to be zero. The error term, u_{t+1} is distributed multivariate normal such that $u_{t+1} \sim N_n(\mathbf{0}, I)$. In a single factor case, f_{t+1} will take the GARCH(1,1) form:

$$\begin{aligned} f_{t+1} &= \sigma_t \varepsilon_{t+1} \\ \sigma_t^2 &= \omega + \alpha f_t^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{2.21}$$

where $\theta = [\omega, \alpha, \beta]$. This volatility specification is consistent with the SR-SARV(1) structure put forth in equation (2.18), where $\gamma = \alpha + \beta$.

The choice of the loading vector Λ and the parameters θ for the factor(s) f_{t+1} must be guided by four key properties. 1) The parameter vector θ must be realistic. 2) Moments of at least order four must exist for the asset returns. 3) The parameter vector θ must be chosen

such that the ARCH LM test can detect the presence of conditional heteroscedasticity within the factor. 4) The loading vector Λ must be chosen such that the ARCH LM test will also be able to detect the presence of conditional heteroscedasticity in the return series. Properties three and four ensure that the ARCH LM test is a useful diagnostic tool.

Choosing θ values that are realistic is straightforward. My application in this paper focuses on daily stock returns. Fitting GARCH(1,1) specifications to a variety of daily U.S. equities yields a range of θ values depicted by the enclosed box in Figure 2.7.

The need for finite fourth moments comes from the structure of the moment conditions. Equation (2.4) indicates that the first moment condition could be written as $(y_t^{(1)})^2[(y_{t+1}^{(2)} - By_{t+1}^{(1)})y_{t+1}^{(1)} - D_1]$. Expanding out this multiplication reveals something like y^4. The existence conditions detailed by Bollerslev (1986) can be used as a tool to determine the existence of the fourth moments of the simulated return series[8]. All points to the left of the frontier in Figure 2.7 represent (α, β) combinations that yield finite 4th moments.

The third key property of the simulations is that θ is chosen such that the ARCH LM test is able to detect the conditional heteroscedasticity present in the factor(s). I investigate the power of the ARCH(1) LM test at detecting the conditional heteroscedasticity in a GARCH(1,1) process across a variety of α and β combinations. The ω is set equal to $1 - \alpha - \beta$ so as to generate unit (unconditional) variance. The graph on the left of Figure 2, depicts a power surface of the LM test for ARCH(1) given a GARCH(1,1) alternative for a sample size of 750 over 100 replications. The nominal size for the tests is 10%. I examine only those (α, β) combinations that are both realistic and correspond to finite fourth order moments. The right hand graph in Figure 2.7 depicts a power surface of the LM test for ARCH(2) errors. In either case, power appears to be highest at low levels of α.

The final consideration for the simulations is that the loading vector Λ be set such that the ARCH LM test be able to detect the conditional heteroscedasticity within the return series. Recall the variance decomposition detailed in equation (2.1). For asset one this equation

[8]The existence conditions outlined in Bollerslev (1986) are derived for a Gaussian process. If the kurtosis is a value other than three, the conditions must be altered. I adopt the Gaussian structure here for illustrative purposes only. This assumption is not maintained throughout the paper.

reduces to $\sigma_1^2 = \lambda_1^2 Var(f) + Var(u_1)$. Recall as well that $Var(f) = Var(u) = 1$ by design. So, the proportion of the total variance σ_1^2 accounted for by the factor is $\frac{\lambda_1^2}{\lambda_1^2+1}$. As λ_1 increases so does that amount of variation in the asset accounted for by the conditionally heteroscedastic factor.

I run a simple Monte Carlo experiment to determine the appropriate Λ in the simulations. I build an ARCH(1) factor as follows:

$$f_{t+1} = \sigma_t \varepsilon_{t+1}$$
$$\sigma_t^2 = \omega + \alpha f_{t-1}^2$$

where $\varepsilon \sim N(0,1)$. I choose ω & α by modeling the sector returns used in the empirical application section of this paper as ARCH(1) processes. Setting $\omega = 0.01$ and $\alpha = 0.50$ seem to be reasonable approximations of this data, which I will detail later. I then construct a series $Y_t = \lambda_i f_t + u_t$ of length $T = 25,000$, where u_t is a standard normal random variable and $\lambda_i \epsilon \Lambda = \{0, 0.1, 0.2, ..., 5\}$. For each λ_i I conduct an LM test for the presence of ARCH(1) effects in the residuals $\widehat{u}_t$. Figure 2.7 plots the p-values from this test against Λ. A λ of at least 2.5 is required for the ARCH LM test to detect the conditional heteroscedasticity in Y.

2.4.2 Gauging the Near Singularity

To illustrate the aforementioned near singularity in the variance/covariance matrix of the moments, consider a small-scale five asset example. I generate 100 paths simulated from equations (2.20) & (2.21) , where $\theta = (0.03, 0.06, 0.91)$ and $\Lambda = [10, 9, 8, 7, 6]'$. Estimating the first step of the sequential testing procedure, that which is associated with H_o^1, is straightforward. However, proceeding to the second step, that which is associated with H_o^2, is challenging.

Define a $(j \times 1)$ vector of instruments $z_t = [1 \ (y_t^{(1)})^2]$. Form the moments $m_{t+1} = z_t \otimes vec(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})y'_{t+1} - D$ and denote their mean over the T observations as g.

Without loss of generality, I set $y^{(1)} = \overline{y}$ for this example. Similar results are found for $\overline{y} = y^{(2)}$ or $y^{(3)}$. Use the identity matrix as an initial weighting matrix and form the typical GMM objective function $J = Tg'Wg$. Denote $S = \frac{1}{T}\sum_{t=1}^{T} m_t m_t'$ as the empirical counterpart of the asymptotic variance matrix of the moments. As per Hansen (1982), the

optimal weighting matrix is the inverse of the asymptotic variance matrix, $W = S^{-1}$.

A singularity among the moments clearly would impede the use of the optimal weighting matrix. Table 2.7 illustrates the condition number of S averaged over 100 sample paths. The Table is separated into three panels, each corresponding to a different sample size $T = \{500, 1000, 2500\}$. The first column indicates the number of GMM iterations used in calculating S. The columns labeled "Cond(S)" capture the condition number of S[9]. Please note that I disregard all paths for which the condition number is infinite in computing this average. The columns labeled "% Fail" capture the number of the 100 sample paths for which S is singular, and thereby precludes GMM estimation.

Consistent, albeit inefficient, estimation of the model is possible via a single GMM iteration in all of the simulated paths. However, attempting to proceed beyond one GMM iteration is difficult. For two GMM iterations, the average condition number of the variance matrix is quite large for the case where the sample size is 500, (1.05e+19). Moreover, approximately 64% of the simulated paths generate variance matrices (S) that are singular. Notice, that this does not appear to be a small sample issue, since large condition numbers persist even as I increase the sample size. In addition, my research suggests that the magnitude of the problem increases as I consider H_o^3 and H_o^4.

In Figure 2.7 I calibrate the Tikhonov Factor for H_o^3 for a given sample path from the example above. The solid red line is the chosen tolerance level of 0.01. The starred blue line is the norm of the difference in the parameter vector. In this example, an α^* of about 1.5 is appropriate.

2.4.3 Determining the Number of Factors

I generate N sample paths of size T for the return series y_t as given in equation (2.20), where $n = 5$ and $K = 1$. The parameters for the variance process detailed in equation (2.21) are (ω, α, β) = $(0.03, 0.06, 0.91)$.

[9]Condition Number of matrix A is computed as $cond(A) = \| A^{-1} \| \cdot \| A \|$, where I use the frobenius norm.

I use a "'regularized"' HAC variance matrix and an instrument vector $z_t = [1\ y_{1,t}^2]'$[10].The loading vectors used vary according to whether size or power of the test is the object of interest.

Determining the empirical size is straightforward. I generate one factor and consider the following loading vector: $\Lambda_S^{(a)} = (10 \quad 9 \quad 8 \quad 7 \quad 6)'$ In this way, the assets differ primarily by their weighting on the factors.

My sequence of hypotheses are as follows:

$$
\begin{aligned}
&H_o^1 : \text{1 factor is sufficient for assets } (y^{(1)}, y^{(2)}) \\
&H_a^1 : \text{1 factor is not sufficient for assets } (y^{(1)}, y^{(2)}) \\
&\vdots \\
&H_o^4 : \text{1 factor is sufficient for assets } (y^{(1)}, y^{(2)}, \ldots, y^{(5)}) \\
&H_a^4 : \text{1 factor is not sufficient for assets } (y^{(1)}, y^{(2)}, \ldots, y^{(5)})
\end{aligned}
$$

If I reject the null at any stage along the sequence, it is considered a failure. I then count the number of failures as a proportion of the number of simulated paths to yield a measure of empirical size.

I test power against several different alternatives, each containing a different number of GARCH factors. Initially, I consider an alternative where the returns are built from two GARCH(1,1) factors. The first factor is as detailed above. The second factor takes a similar form, except that $(\omega, \alpha, \beta) = (0.05, 0.10, 0.85)$. The factors have zero correlation by design. The loading vector $\Lambda_P^a = [\Lambda_S^a\ \Lambda_S^a] \odot [I_2\ \iota_{2,3}]'$, where I_2 is a (2 x 2) identity matrix and $\iota_{2,3}$ is an (2 x 3) matrix of ones.[11]

The shape of this loading vector is useful for identification purposes, and follows directly from defining $B = \overline{\Lambda}\overline{\Lambda}^{-1}$ as in section 2.3.1. Also, as discussed in section 2.4.1, the magnitude of the elements of the loading vector correspond to the amount of variation the factors are able to capture.

As when testing empirical size, a rejection of the null anywhere along the sequence of tests

[10]Although a Newey-West variance matrix should not be necessary since the moment conditions are given by martingale difference sequences, using a HAC estimator is often a prudent strategy in finite samples.

[11]The symbol $\odot$ is used to represent the Hadamard product.

is considered a failure. The probability of rejection is then measured as the number of failures as a proportion of the number of simulated paths.

The base case for the balance of this exercise involves simulating return paths via equation (2.2) with the following: (number of simulated paths) $N = 250$, (sample size) $T = 1000$, $z_t = [1 \ y_{1,t}^2]$, $\overline{y} = y^{(1)}$, loading vector Λ^a, and weighting matrix equal to the inverse of the regularized HAC variance matrix. For a single path I conduct an ARCH(1) LM test and gather the p-values for each portfolio. I repeat this test for all 250 sample paths. The first column of Table 2.7 records the p-values averaged across the 250 paths. The small p-values suggest I can reject the null of No ARCH effects for all 5 portfolios. In addition, Figure 2.7 illustrates the size and power of H_0^4 versus a two factor alternative. Notice that the test has only slight size distortions and has excellent levels of power.

These findings are robust to various characterizations of the model. Table 2.7 details the empirical size and power of H_0^4 as I alter the model relative to the base case. All calculations are shown for a 10% nominal size. I summarize the results here.

The first panel depicts a reasonably sized test when I use $\overline{y} \epsilon \{y^{(1)}, y^{(2)}, y^{(3)}\}$. However, there are size distortions when using $\overline{y} \epsilon \{y^{(4)}, y^{(5)}\}$. In all cases, the test is powerful.

The second panel of Table 2.7 suggests that there are slight size distortions when the sample grows large, yet the test is quite powerful in all cases considered.

The third panel indicates that both empirical size and power increase as the magnitude of the loading vector Λ increases.

The fourth panel of Table 2.7 suggests that the power increases from an already high level as the alternative hypothesis considered expands from a system constructed from 2 factors to 3 or 4 factors.

The last panel varies the choice and number of instruments used in estimation. In each of the first five rows, the instruments are a vector of ones appended by a one period lagged value of the square of each asset's returns. The last row of the panel captures the case where the instrument vector includes a vector of ones appended by the average of the one period lagged values of squared returns for all five assets. Specifically, $z_t = \left[1 \ \frac{1}{5} \sum_{j=1}^{5} (y_t^{(j)})^2\right]'$. In general, size and power are quite robust to instrument choice.

For each simulated path I form the time series associated with four auxiliary portfolios of the form $(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})$. As suggested earlier, these portfolios should be conditionally homoscedastic. I confirm this hypothesis by searching for ARCH(1) effects via an LM test. The p-values from the test are averaged over the 250 sample paths for each of the portfolios. The first column of Table 2.7 captures the p-values from the ARCH LM test for the five base assets, while the second column captures the p-values for the four auxiliary portfolios. The typical p-value rises from about 0.10 in the base assets to about 0.50 among the auxiliary portfolios, indicating that no discernible ARCH(1) effects remain in the auxiliary portfolios.

In addition, I examine each of the auxiliary portfolios for DCC effects. The first column of Table 2.7 captures the p-values from the DCC test for the five base assets. The second column of the Table captures the p-value from the DCC test for the four auxiliary portfolios. The first row of the Table pertains to the full sample of 250 simulations. The p-values listed are averaged over the 250 simulated paths. For the full sample, the p-value for the auxiliary portfolios is about 0.09 percentage points higher (0.733 vs. 0.644) for the auxiliary portfolios than for the base assets, which implies that the auxiliary portfolios are able to alleviate some of the time variation in the correlation matrix of the assets. However, this is not a difficult feat since there wasn't strong evidence of DCC effects in the base assets. To gauge better the ability of this procedure to accommodate the dynamics of the five asset system, I examine only those simulated paths for which DCC effects are strong. The second row of Table 2.7 pertains only to the 14 simulated paths for which the p-value is less than 0.15. The average p-value of 0.094 suggests strong DCC effects among the base assets. The auxiliary portfolios pertaining to these 14 paths has an average p-value of 0.641, thereby eliminating the DCC effects. When DCC effects are present, Phase 1 is able to accommodate for the time varying correlation matrix.

I conduct a second exercise to determine how well the model can identify the number of latent factors in a system of asset returns. Once again consider the base case simulation. Each row of Table 2.7 represents the first element used to define $\overline{y}$. Each column indicates the number of factors determined by the sequential testing procedure. The entries indicate the frequency of simulated paths for which Phase 1 determined a particular number of factors. For

instance, row one / column one indicates that only 83 out of the 100 simulated paths suggest that one factor is sufficient to accommodate the conditional heteroscedasticity present in the five asset system while using $y^{(1)}$ as the first column of $\overline{y}_{t+1}$. Table 2.7 is identical in structure except that the true simulated series is built from two factors rather than one. Specifically, the assets are built from two GARCH(1,1) factors with parameters $(\omega, \alpha, \beta) = (0.03, 0.06, 0.91)$ and $(0.05, 0.10, 0.85)$, where I use a loading vector $\Lambda = [\Lambda_S^a \ \Lambda_S^a] \odot [I_2 \ \iota_{2,3}]'$.

Table 2.7 indicates that a vast majority of the simulations identify correctly the number of factors in the one-factor system. In addition, the model rarely misidentifies this one factor system for two factors. However, three to four factors are chosen mistakenly in $(15 - 20\%)$ of the simulations. Table 2.7 suggests that very seldom does Phase 1 underestimate the number of factors in the two-factor system. For example, when $\overline{y} = y_1$, only 2% of the simulated paths chose a single factor as sufficient. Setting $\overline{y} = y^{(1)}$ or $y^{(2)}$, the two most conditionally heteroscedastic assets by definition, seems to work best in the simulations. 98% (97%) of the simulations correctly pick a single factor when setting $\overline{y} = y^{(1)}$ $(y^{(2)})$. However, as the choice of $\overline{y}$ moves to less conditionally heteroscedastic assets, the number of factors tends to be overestimated. For example, when $\overline{y} = y^{(3)}$, 34% of the simulations incorrectly pick three factors and 2% pick four factors. These findings are robust to instrument choice. Although not shown here, as I vary the instrument vector, the accuracy of Phase 1 remains consistent with the results reported.

2.4.4 Full Model Estimation

This section examines the efficacy of Phase 2 of the estimation procedure through a Monte Carlo - type analysis. I extract the latent factors implied by the model and compare these to the true simulated factors. With the factors now 'observed', I can model them via conventional time-varying volatility techniques such as GARCH or SV models, and compute conditional forecasts. This opens the door to a Markowitz-style portfolio optimization, where I can create tracking portfolios of the dynamic allocation strategy implied by the MVFSV model. I utilize this tracking portfolio approach in the Empirical Application section of this paper.

I begin the evaluation by simulating a system of five asset returns corresponding to our

base case from the previous section. The persistence parameters associated with the single common factor are $(\omega, \alpha, \beta) = (0.03, 0.06, 0.91)$, with loading vector $\Lambda^a = [10, 9, 8, 7, 6]$.

In order to judge Phase 2 fairly, I isolate it from any potential errors that may arise during Phase 1. I accomplish this by assuming that the number of factors is given, thereby avoiding Phase 1 entirely.

This single factor structure yields the following set of moment conditions for a model with a constant risk premium:

$$\begin{aligned} vecE_t[(\overline{\overline{y}}_{t+1} - B\overline{y}_{t+1})y'_{t+1}] &= vec(\overline{\overline{\widehat{\mu}}} - B\overline{\widehat{\mu}})\widehat{\mu}' + (\Omega_2 - B\Omega_1) \\ diagE_{t-1}[(1-\gamma L)(y_{t+1}y'_{t+1})] &= diag(\widehat{\mu}\widehat{\mu}' + \widehat{\Lambda\Lambda' + \Omega})(1-\gamma_1) \end{aligned} \tag{2.22}$$

where $\widehat{\mu}$ and $\widehat{\Lambda\Lambda' + \Omega}$ are the sample counterparts of the mean and variance / covariance matrix of returns, respectively.

Evaluating the accuracy of parameters associated with the loading vectors (Λ), the factor volatilities (γ), and the factors themselves (f_t), requires some work.

Recall the conditional variance specification:

$$\Sigma_t = \Lambda D_t \Lambda' + \Omega \tag{2.23}$$

and notice that we can characterize the loading vector as $\Lambda = [I \ B]'\overline{\Lambda}$. Also recall that while B is identified by the model, Λ is not. In fact, only the range of Λ is identified. This is due primarily to the possibility of a non-diagonal variance of the errors, Ω. This non-diagonality implies that any constant part of the conditional variance of the factors can be transferred to the variance of the errors by a simple rescaling of the loading vectors.

For the purposes of evaluating the model, I take advantage of this re-scaling property by normalizing $\overline{\Lambda}$, without any identifiable impact on Σ_t. The normalization scheme I use exploits the link between the factors and their loading vectors as well as the fact that the factors have unit unconditional variance by design.

The normalized value of the loading vector is computed as $\Lambda = [I \ \widehat{B}]'\overline{\Lambda}$, where $\widehat{B}$ is the GMM estimate. Moreover, the normalized error variance is then easily computed as

$\widehat{\Omega} = \widehat{\Lambda\Lambda' + \Omega} - \widehat{\Lambda}\widehat{\Lambda}'$.

For ease of exposition, I set $\overline{\Lambda} = I_k \times a$, where a is some constant. For a given a, we can normalize the loading vector $\widehat{\Lambda}(a)$ and extract the latent factors via simple OLS regressions of $y_t - \widehat{\mu}$ on $\widehat{\Lambda}(a)$ for each period $t = 1, ..., T$. The factor takes the usual form $\widehat{f}(a)_t = (\widehat{\Lambda}(a)'\widehat{\Lambda}(a))^{-1}\widehat{\Lambda}(a)'(y_t - \widehat{\mu})$. I choose a to minimize the distance between the sample variance of the factors and the population expectation of 1. In other words, $a = argmin_{a \in A}||(\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(a)_t - \overline{\widehat{f}}(a))^2 - 1||^2$.

Panel A of Tables 2.7 and 2.7 illustrate the estimated value of the loading vector averaged over 250 paths for $T = 500$ and $T = 1,000$, respectively. The column labeled 'Model' captures estimates from the MVFSV model, and the column label 'Simulated' captures the values actually simulated. The estimates of the loading vector are quite accurate, both in terms of order of magnitude and shape; i.e. the first element is the largest and the last element is the smallest.

Panel B of Tables 2.7 and 2.7 compares the unconditional moments of the simulated factors with the extracted factors after normalization. Regardless of sample size, the first four sample moments of the extracted factors match remarkably well those of the simulated factors. For instance, the kurtosis of the extracted factor is 3.250 averaged over the 250 paths, which matches precisely the kurtosis of the simulated factors. Moreover, the correlation between the extracted and simulated factors is above 0.99 in both sample sizes.

Panel C of Tables and 2.7 and 2.7 compare the conditional behavior of the extracted factors. Given the SR-SRAV(1) specification of the volatility dynamics, we know $\gamma = \alpha + \beta$ from a GARCH(1,1) model. The simulated values of $\alpha = 0.062$ and $\beta = 0.822$ imply a simulated value of $\gamma = 0.884$, for the case of $T = 500$[12]. The estimates of α & β are quite good; 0.063 and 0.827 respectively. However, the estimates of γ are somewhat biased[13]. For $T = 500$, $\widehat{\gamma} = 1.5$. Note that the estimates improve as the sample size increases. For $T = 1,000$, $\widehat{\gamma} = 1.3$. Not

[12]Notice that the values of α & β actually simulated do not match precisely the "true" values $(0.06, 0.91)$. The difference is due to finite sample error during the simulation procedure. After simulation, I estimate the factor processes with a GARCH(1,1) model in order to obtain more accurate values of α and β.

[13]The estimate of γ reported here is that which is found from Phase 2, not the summation of $\hat{\alpha}$ and $\hat{\beta}$.

shown in the Tables are results for $T = 2,500$ ($\widehat{\gamma} = 0.991$) and for $T = 10,000$ ($\widehat{\gamma} = 0.993$)

Also of interest is the internal consistency of the estimates. In particular, Phase 2 yields an estimate of B and of the matrix $\Omega_2 - B\Omega_1$, where Ω_1 is the upper $(k \times n)$ submatrix of Ω and Ω_2 is the lower $(n-k \times n)$ submatrix of Ω. Normalizing the loading vector allows me to impute an estimate of Ω via $\widehat{\Lambda'\Lambda + \Omega} = \widetilde{\Omega}$. If the estimation procedure is internally consistent, then $\widehat{\Omega_2} - \widehat{B}\widehat{\Omega_1}$ should equal $\widehat{\Omega_2 - B\Omega_1}$.

The simulation results suggest that the model is indeed internally consistent. The square of the element-wise difference in the estimators provides a measure of comparison of these estimators, akin to MSE computed as $[vec(\widehat{\Omega_2} - \widehat{B}\widehat{\Omega_1}) - \widehat{\Omega_2 - B\Omega})]' * [vec(\widehat{\Omega_2} - \widehat{B}\widehat{\Omega_1}) - \widehat{\Omega_2 - B\Omega})]/[(n-k) \times n]$. For $T = 500$, this element-wise difference is only 0.8. As the sample size increases, this measure drops to 0.13, 0.005, and 0.001 for $T = 1,000$, $T = 2,500$, and $T = 10,000$, respectively.

2.5 Empirical Application

In this section I use the MVFSV model to investigate the dynamics of the U.S. equity market. I examine portfolios representing the 12 sectors of the U.S. economy. I determine how many factors are required to accommodate the conditional heteroscedasticity in this system of returns. I then use the MVFSV model to forecast the conditional variance matrix of returns. These forecasts pave the way for a dynamic asset allocation strategy. I will track the investment strategy and compare its performance to a strategy implied by the DCC model.

2.5.1 Data

I segment the CRSP universe of U.S. equities into twelve value weighted sector portfolios for every trading day from 01/02/1990 through 12/31/2006. The portfolios are formed via a time series of the two-digit SIC codes obtained from Ken French's website. Table 2.7 provides summary statistics, illustrating the non-Gaussianity, negative skewness and excess kurtosis typical of daily stock returns.

Figures 2.7, 2.7, 2.7 depict the cumulative returns of each of the 12 sectors over the 17 year

period. Each graph illustrates the growth of a single dollar invested on 01/02/1990 and held for the entire period.

The dynamic allocation strategies I will detail allow for the presence of a risk free rate. I follow French, Schwert and Stambaugh (1987) by using 1-month U.S. Treasury yields as a standard proxy. Specifically, monthly yields are gathered from the average of bid and ask prices for the U.S. government security that matures closest to the end of the month. Daily yields are then computed by dividing the monthly yield by the number of trading days in the month.

2.5.2 Estimation & Results

The Empirical Application's design centers around an investor who trusts the dynamic nature of their model. Each month the investor estimates the model and generates one month's worth (approximately 21 trading days) of conditional variance forecasts. With these forecasts, the investor solves a portfolio optimization problem and re-balances their portfolio every trading day. The investor uses three years (approximately 756 trading days) worth of past returns data for estimation.

The investor relies upon the sequential testing procedure of Phase 1 to determine the proper ordering of the sector portfolios. Twelve choices are available to begin the sequence. The investor examines each choice and selects the ordering that achieves the best balance between the size of the factor structure and the model's ability to accommodate the conditional heteroscedasticity in the returns. Specifically, the investor chooses the smallest factor representation that has auxiliary portfolios that are on average less conditionally heteroscedastic than the original sectors, as measured by individual LM tests, and exhibit less dynamic conditional correlation, as measured by a joint test for constant conditional correlation.

The first estimation period runs from 01/02/1990 through 01/04/1993. I estimate Phase 1, pick the ordering of the sectors that meets the selection rule described above, form the auxiliary portfolios, and test them for ARCH and DCC effects. I then roll the investor's estimation window forward one month, using returns from 02/01/1990 through 02/01/1993. I re-estimate the model and form new auxiliary portfolios. This process continues for a total of

168 times until the end of the last estimation period is 12/31/2006.

Figure 2.7 depicts the number of factors selected by the investor at each of the 168 estimation points. As intuition would suggest, the relative tranquility of the economy and market in the mid 1990's generally required only a single factor. However, the turbulent times of late 90's and early 2000's required as many as seven factors. It is in times such as these that a factor analytic approach might not be useful as it removes the key advantage of parsimony in the model.

The auxiliary portfolios formed from these factor structures are able to accommodate the individual ARCH effects among the assets rather well. Fig 2.7 illustrates the average p-values from ARCH(1) tests. The solid line is the average of the 12 p-values from the LM test on the sector portfolios. The stars are the average of the (n-k) p-values implied by the auxiliary portfolios. The null on the LM test for ARCH(1) effects is that no ARCH is present. Therefore, a low p-value is suggestive of strong ARCH effects, while a large p-value is suggestive of conditional homoscedasticity. The relative tranquil period of the mid-1990's, when only a single factor was needed, was able to accommodate the ARCH effects quite well. In 1995 for instance, the average p-value for the sector portfolios was approximately zero. Yet, the average p-value of the auxiliary portfolios was as high as 0.55.

Fig 2.7 illustrates the auxiliary portfolios' ability to accommodate the joint dynamics in the data. The solid line represents the p-value from a test for constant conditional correlation for the twelve sector portfolios. The stars again represents the p-value for (n-k) auxiliary portfolios. The null hypothesis of this test is that the returns exhibit constant conditional correlation. Therefore, a low average p-value is indicative of DCC effects. The MVFSV model is unable to accommodate for the strong DCC effects during the early 1990's. However, in the mid-1990's DCC effects are handled quite well. The model's ability to generate auxiliary portfolios with constant conditional correlation is somewhat erratic during the late 1990's boom. And interestingly, there is little evidence of DCC effects in the sector portfolios during the 2003-2006 period.

As described in the Simulation section, the estimates of B allow us to normalize the loading vector, extract the factors, and forecast the conditional variance matrix of returns. Forecasts

of the mean of returns, on the other hand, are computed simply as the sample average over the estimation period. As such, the investor can be characterized as a volatility-only forecaster. With these forecasts, the investor conducts a Markowitz-style portfolio optimization. Their objective function is to minimize volatility of the portfolio subject to a 10% target rate of return. I allow for short sales as well as for a time varying risk free rate, proxied for by the 1-month U.S. Treasury bill. Details of the investor's objective functions and the analytic solution for the portfolio weights are as per Han (2006) and are found in the Appendix.

I compare the portfolio implied by the MVFSV model against two portfolios. First, is a Buy & Hold benchmark consisting of the return on a value-weighted portfolio of the entire CRSP universe. Second, I consider a portfolio formed from forecasts made by the DCC model. The MVFSV and DCC models are referred to as dynamic strategies since they involve dynamically updating the portfolio each trading day.

Several criteria are used to evaluate the performance of the Buy & Hold and dynamic strategies. μ_p is the average daily return of the portfolio. σ_p is the standard deviation of the daily portfolio returns. r_f is a measure of the risk free rate. $SR = (\mu_p - r_f)/\sigma_p$ is an ex-post Sharpe Ratio of portfolio returns. The Sharpe Ratio, however, is known to underestimate the performance of dynamic strategies because it does not account for time variation in conditional volatility of the portfolio. Therefore, I also consider $M2 = \sigma_b(SR_p - SR_b)$, where the b subscript indicates the static Buy & Hold benchmark. $M2$ can be seen as a risk adjusted measure of abnormal return. $IR = \frac{\mu_p - \mu_b}{\sigma_p - \sigma_b}$, is the Information Ratio. A higher Information Ratio is indicative of better performance relative to the benchmark. Z is the percentage of trading days the dynamic strategy has a higher return than the benchmark strategy, providing an indication as to whether out-performance is due to a few large outliers. Finally, ϕ is the performance-fee measure of Fleming, Kirby and Ostdiek (2001). It measures the hypothetical fee that an investor would be willing to pay to switch from the benchmark static strategy to a given dynamic strategy, without being worse off. The performance fee ϕ solves:

$$\frac{1}{T}\sum_{t=0}^{T-1}[(R_{d,t+1} - \phi) - \frac{\gamma}{2(1+\gamma)}(R_{d,t+1} - \phi)^2] = \frac{1}{T}\sum_{t=0} T - 1[R_{b,t+1} - \frac{\gamma}{2(1+\gamma)}R_{b,t+1}^2]$$

where $R_{d,t+1}$ and $R_{b,t+1}$ are the realized gross returns of the dynamic and static benchmark strategies, respectively.

I consider trading costs as per Marquering and Verbeek (2004), where costs are equal to τ percent of the value traded. Total transactions costs are captured by $\tau W_t|\Delta\omega_{t+1}|$, where W_t denotes wealth at time t, and $\Delta\omega_{t+1} = \omega_{t+1} - \omega_t$. Returns after transaction costs are measured by $y_{p,t+1} - \tau|\Delta\omega_{t+1}|$. I follow Marquering and Verbeek (2004) by considering low, medium, and high transaction costs of 0.1%, 0.5%, and 1.0% of the value traded. The Buy & Hold benchmark portfolio is assumed to have no transaction costs.

Figures 2.7 through 2.7 illustrate the performance of the dynamic allocation strategies for the minimum volatility investor at no, and a medium level of transaction costs, respectively. I omit the low and high transactions cost cases to save space. The findings are similar. Four criteria are selected to summarize succinctly the relative performance of the portfolios[14]. The upper left panel in each figure contains $\overline{u}_p$, the daily return averaged over the roughly 252 trading days in each year from 1993 through 2006. Units are in percent, so that 0.1 is 0.1% per day. The top right panel illustrates the average daily standard deviations σ_p. The bottom left and right illustrate the M2 and ϕ measures, respectively.

Focus first on Figure 2.7. In terms of absolute daily returns, the performance of the dynamic strategies relative to the buy & hold benchmark can change dramatically. However, a clear pattern emerges by examining the relative ranking among the two dynamic strategies. From 1993 through 1997, the portfolio implied by the DCC model outperforms that implied by the MVFSV model. From 1998 through 2002, the performance flips in favor of MVFSV. The 2003-2006 period again flips in favor of the DCC model.

The MVFSV portfolio has a higher level of daily standard deviation than the DCC portfolio, which contributes to the DCC's relatively high M2 for much of the sample period. However, despite this variability in returns, the MVFSV portfolio is still able to generate better M2 performance during most of the late 1990's boom period. A similar finding holds for the Φ measure of performance.

[14]The general theme reported here does not change as we consider the other performance criteria mentioned.

The MVFSV model's performance, however, is accompanied by a large amount of turnover in the portfolio. To illustrate, consider a measure of total portfolio activity (TPA) as follows: $TPA = \sum_{i=1}^{n} |w_{i,t+1} - w_{i,t}|$, where n is the number of assets in the portfolio, and w are the asset weights. Figure 2.7 illustrates the total portfolio activity for the MVFSV and DCC models, smoothed over a rolling three month period for illustrative purposes.

For a majority of the sample period, the MVFSV and DCC models exhibit roughly equal levels of portfolio activity. However, the MVFSV portfolio experiences periodic surges in the amount of transactions. Such high levels of activity clearly will impact portfolio performance when transactions costs are considered. For example, at medium levels of transactions costs, as seen in Figure 2.7, the MVFSV under-performs the DCC model in almost all metrics considered.

In summary, the MVFSV model generates portfolios that perform comparably to the DCC model according to several criteria. Trading activity, however, is important. In the presence of transaction costs, significant levels of portfolio activity could have serious effects on absolute performance and cause the DCC model to outperform the MVFSV on a relative basis.

Also of interest is the model's ability to estimate the conditional covariance matrix of returns. Such an assessment may not only be useful in explaining the relative and absolute performance of the MVFSV strategy, but may also be useful in option pricing exercises and risk management.

I gauge the accuracy of the covariance matrix estimation by using a measure of realized covariance. Following French, Schwert and Stambaugh (1987), I use the sum of daily cross-returns as a measure of the monthly realized covariance. Specifically, $RCov_t = \sum_{\tau=1}^{M} y_{i,\tau} y_{j,\tau}$ $\forall i, j$, where M is the number of days in month (t). Similarly, I compute the implied monthly covariances of the MVFSV and DCC strategies using the forecasted daily covariances from each model. I then calculate the absolute estimation error for each of the 78 unique elements of the conditional covariance matrix as $|RCov_t^{MVFSV} - RCov_t|$. The estimation error for the DCC model is computed analogously.

I compare the models' ability to forecast the conditional covariance matrix by examining the error of the MVFSV model relative to that of the DCC model. Specifically, I compute the

element-by-element matrix division: $|RCov_t^{MVFSV} - RCov_t|/|RCov_t^{DCC} - RCov_t|$. Values of this relative error less than one suggest that the MVFSV model is favorable.

By averaging the relative errors over the 78 unique elements of the covariance matrix, I find that the MVFSV model is able to outperform the DCC model in only about 30% of the sample.

Looking at each element of the covariance matrix separately, however, reveals a much more diverse picture of relative performance. For example, Figure 2.7 depicts the relative error of the covariance between the Business Equipment and Durable Goods sectors. The MVFSV outperforms DCC at almost every point in the sample. Yet, Figure 2.7 suggests that the MVFSV model is not as accurate as the DCC model at explaining the variance of the Business Equipment sector, which experienced the largest price movements during the sample.

Whereas the DCC model does a good job at forecasting variances, the MVFSV model may have an advantage at forecasting covariances. A further examination of these findings is left as future work.

2.6 Conclusion

In this paper I illustrate that the multivariate latent factor stochastic volatility model (MVFSV) of DR can be implemented successfully, and yield potentially significant investment implications. Tracking portfolios of dynamic asset allocation strategies reveal that the MVFSV model performs comparably with the DCC model. Yet, the MVFSV model does so while offering a parsimonious way to capture the joint dynamics of a system of asset returns without resorting to a full parameterization of conditional probability distributions. The issue of comparing the statistical fit of this MVFSV model to other factor stochastic volatility models remains unanswered.

An obvious extension of this paper is to include a time varying risk premium in Phase 2, as outlined by DR. This may improve the tracking portfolio performance as well as open the door to testing the price of idiosyncratic risk as in King, Sentana and Wadhwani (1994). Moreover, as already noted by DR, a GMM approach also is possible when allowing for time variation in

the diagonal elements of the residual variance matrix Ω. This extension should be relevant for those assets whose heteroscedasticity cannot be captured fully by a few common factors.

2.7 Tables & Figures

Table 2.1: Illustrating the Near Singularity

Cond(S) is the condition number of the variance matrix of moments. %Fail is the proportion of simulated paths for which the GMM optimization fails to converge due to a singularity problem among the moments.

Singularity of the GMM variance matrix in H_o^2						
	T=500		T=1000		T=2500	
GMM It.	Cond(S)	% Fail	Cond(S)	% Fail	Cond(S)	% Fail
1	2.94e+19	0%	1.59e+19	0%	2.84e+19	0%
2	1.05e+19	64%	7.12e+18	69%	3.67e+18	75%

Table 2.2: Determining Size and Power Of Phase 1 Tests

Testing the empirical size and power of the overidentified restrictions test associated with the search for the number of common factors. The analysis pertains to a simulated system of five asset returns built from a single GARCH(1,1) factor. The base case for the study is $\overline{y} = y^{(1)}$, *T=1000,* $\Lambda = \Lambda^a$, *and* $z_t = [1\ (y_t^{(1)})^2]$, H_a *= 2 factors. The top panel alters* $\overline{y}$, *the second panel alters sample size, the third panel alters the loading vectors , the fourth panel alters the number of factors in the alternative hypothesis, and the fifth panel alters the instrument. Nominal size is* 10%.

		Size	**Power**
$\overline{y}$			
	$\overline{y} = y^{(1)}$	0.152	0.952
	$\overline{y} = y^{(2)}$	0.148	0.996
	$\overline{y} = y^{(3)}$	0.164	0.988
	$\overline{y} = y^{(4)}$	0.312	0.980
	$\overline{y} = y^{(5)}$	0.336	0.992
T			
	500	0.124	0.952
	1000	0.152	0.952
	2500	0.252	0.963
Λ			
	$\Lambda^b = [5, 4, 3, 2, 1]$	0.112	0.812
	$\Lambda^a = [10, 9, 8, 7, 6]$	0.152	0.952
	$\Lambda^c = [20, 19, 18, 17, 16]$	0.296	0.996
H_a			
	2 factors	0.152	0.952
	3 factors	-	1.00
	4 factors	-	0.992
z			
	$(y_t^{(1)})^2$	0.152	0.952
	$(y_t^{(2)})^2$	0.104	1.0
	$(y_t^{(3)})^2$	0.144	0.992
	$(y_t^{(4)})^2$	0.144	0.992
	$(y_t^{(5)})^2$	0.128	0.996
	$(y_t^{(j)})^2\ \forall j = 1, .., 5$	1.0	1.0

Table 2.3: Conditionally Homoscedastic Portfolios

P-Values from ARCH(1) LM tests. H_0: No ARCH Effect. Base Assets are five simulated asset returns y. Auxiliary Portfolio are linear combinations of the Base Assets, which take the form $(\overline{\overline{y}} - B\overline{y})$. The p-values are averaged over 250 simulated paths.

	Base Assets	Auxiliary Portfolios
1	0.104	0.518
2	0.103	0.501
3	0.106	0.521
4	0.104	0.404
5	0.121	

Table 2.4: Constant Conditional Correlation Portfolios

P-Values from DCC Test. H_0: Constant Conditional Correlation. Base Assets are five simulated asset returns y. Auxiliary Portfolios are linear combinations of the Base Assets, which take the form $(\overline{\overline{y}} - B\overline{y})$. The p-values are averaged over 250 simulated paths. "Strong DCC" refer only to those simulations for which the p-value < 0.15.

	Base Assets	Auxiliary Portfolios
Full Sample	0.644	0.733
Strong DCC	0.094	0.641

Table 2.5: Recovering a Single Factor

Simulate 100 sample paths of length 1000 for a five asset system. Built from a single GARCH(1,1) factor using loading vector $\Lambda = \Lambda_S^a$, and estimated with $z_t = [1\ (y_t^{(1)})^2]$. Use Phase 1 sequential procedure to determine # of latent factors. Each row indicates the first element of $\overline{y}_{t+1}$. Each column indicates signifies the number of factors suggested. For instance, entry (1,1) suggests that if we set the first element of $\overline{y}_{t+1}$ equal to $y_{t+1}^{(1)}$, then only 83 simulated paths suggest a single common factor is sufficient.

Begin populating $\overline{y}$ with:	# of Factors				
	1	2	3	4	5
$y^{(1)}$	83	0	2	10	5
$y^{(2)}$	84	1	1	13	1
$y^{(3)}$	78	0	1	16	5
$y^{(4)}$	76	0	3	16	5
$y^{(5)}$	67	0	12	15	6

Table 2.6: Recovering Two Factors

Simulate 100 sample paths of length 1000 for a five asset system. Built from two GARCH(1,1) factors using loading vector $\Lambda = [\Lambda_S^a \ \Lambda_S^a] \odot [I_2 \ \iota_{2,3}]'$*, and estimated with* $z_t = [1 \ (y_t^{(1)})^2]$*. Use Phase 1 sequential procedure to determine # of latent factors. Each row indicates the first element of* $\overline{y}_{t+1}$*. Each column indicates signifies the number of factors suggested. For instance, entry (1,1) suggests that if we set the first element of* $\overline{y}_{t+1}$ *equal to* $y_{t+1}^{(1)}$*, then only 2 simulated paths suggest a single common factor is sufficient.*

Begin populating $\overline{y}$ with:	# of Factors				
	1	2	3	4	5
$y^{(1)}$	2	98	0	0	0
$y^{(2)}$	1	97	2	0	0
$y^{(3)}$	2	62	34	2	0
$y^{(4)}$	0	79	20	1	0
$y^{(5)}$	1	82	16	1	0

Table 2.7: Recovering Factors and Loading Vector ($T = 500$)

Base Case simulation with 250 paths, each of length 500. Column labeled 'Model' contains the parameter estimates from Phase 2 of MVFSV. Column labeled 'Simulated' are the parameter values actually simulated. Statistics reported are averaged over all 250 paths. Panels A describes the estimated loading vector. Panels B & C provide estimates of the extracted factors. Panel B lists unconditional moments of the factors, while Panel C explores the conditional moments. Panel C also lists estimates of γ derived from Phase 2.

	Model	**Simulated**
Panel A: Λ		
$y^{(1)}$	10.028	10.0
$y^{(2)}$	8.924	9.0
$y^{(3)}$	7.979	8.0
$y^{(4)}$	7.007	7.0
$y^{(5)}$	6.081	6.0
Panel B: f_t - Unconditional		
Mean	0.000	0.000
Variance	1.000	1.024
Skewness	-0.020	-0.020
Kurtosis	3.250	3.250
Correlation	0.998	-
Panel E: f_t - Conditional		
α	0.063	0.062
β	0.827	0.822
γ	1.526	0.884

Table 2.8: Recovering Factors and Loading Vector ($T = 1000$)

Base Case simulation with 250 paths, each of length 1000. Column labeled 'Model' contains the parameter estimates from Phase 2 of MVFSV. Column labeled 'Simulated' are the parameter values actually simulated. Statistics reported are averaged over all 250 paths. Panels A describes the estimated loading vector. Panels B & C provide estimates of the extracted factors. Panel B lists unconditional moments of the factors, while Panel C explores the conditional moments. Panel C also lists estimates of γ derived from Phase 2.

	Model	**Simulated**
Panel A: Λ		
$y^{(1)}$	9.978	10.0
$y^{(2)}$	9.005	9.0
$y^{(3)}$	8.022	8.0
$y^{(4)}$	6.960	7.0
$y^{(5)}$	6.023	6.0
Panel D: f_t - Unconditional		
Mean	-0.001	0.001
Variance	1.000	1.006
Skewness	-0.016	-0.015
Kurtosis	3.296	3.299
Correlation	0.998	-
Panel E: f_t - Conditional		
α	0.061	0.061
β	0.893	0.893
γ	1.283	0.954

Table 2.9: Descriptive Statistics of 12 Sector Returns

Descriptive Statistics: Daily Returns of 12 sectors in the U.S. market over the period (01/02/1990 - 12/29/2006). Mean and Standard Deviation are annualized and presented in %; i.e. a mean return of 11.83 indicates a 11.83% *annualized return. JB-Pval is the p-value from the Jarque-Bera test for normality, the null hypothesis of which is a Gaussian process.*

	Sector	Mean	Std. Dev.	Skewness	Kurtosis	JB-Pval
1	Non-Durables	11.83	13.73	-0.06	6.58	0.00
2	Durables	9.81	19.98	-0.07	6.68	0.00
3	Manufacturing	13.71	16.25	-0.12	6.18	0.00
4	Energy	14.40	19.68	0.06	5.25	0.00
5	Chemicals	11.92	16.02	-0.47	13.83	0.00
6	Business Equipment	14.94	27.62	0.30	8.24	0.00
7	Telecommunications	7.95	19.13	-0.03	6.87	0.00
8	Utilities	10.79	13.97	-0.30	11.39	0.00
9	Shops	12.26	18.07	0.03	6.98	0.00
10	Health	13.12	18.48	-0.19	6.47	0.00
11	Money	15.75	17.81	0.13	7.04	0.00
12	Other	8.09	16.83	-0.23	9.33	0.00

Figure 2.1: Time Varying Correlations

Equally weighted average of cross-sector correlation coefficients. Computed over preceding three months for 12 sectors of the U.S. economy.

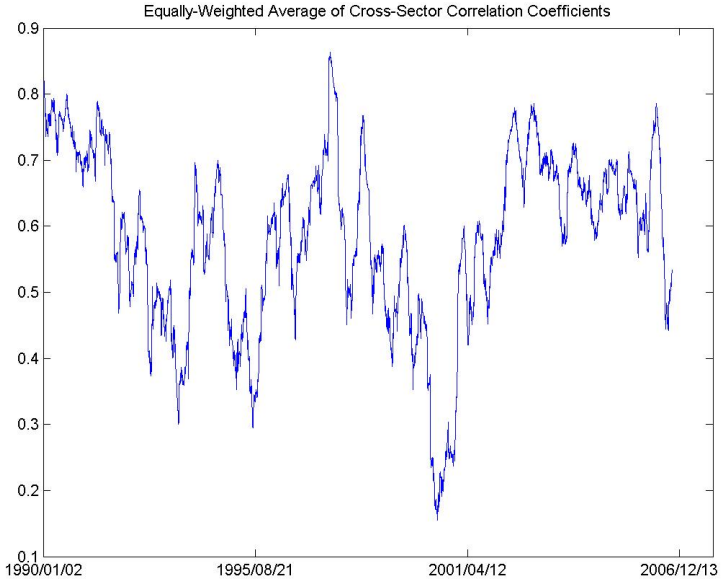


Figure 2.2: Moment Existence

Points to the left of the solid line indicate (α, β) combinations for which fourth moments of returns exist. The box indicates (α, β) combinations that are realistic given daily U.S. equity returns.

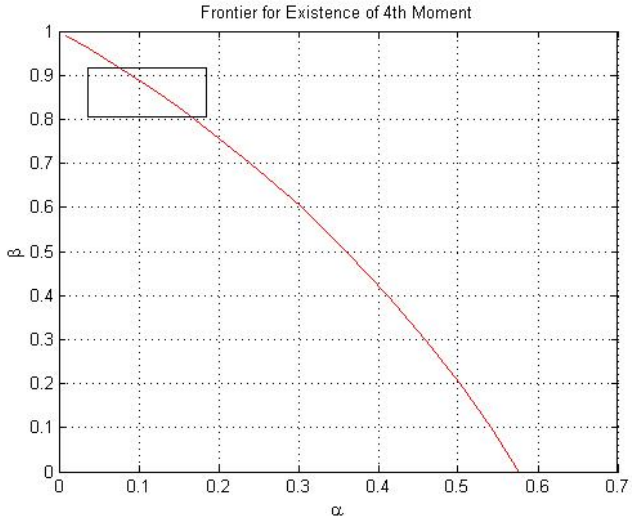


Figure 2.3: Power of ARCH LM Test

Simulation of 100 paths of a 750 observation GARCH(1,1) process for a range of α and β values. I conduct an ARCH LM test and compute power as the proportion of the simulations for which the null of no ARCH effects is rejected. The right hand figure tests for ARCH(2), while the left hand side tests for ARCH(1).

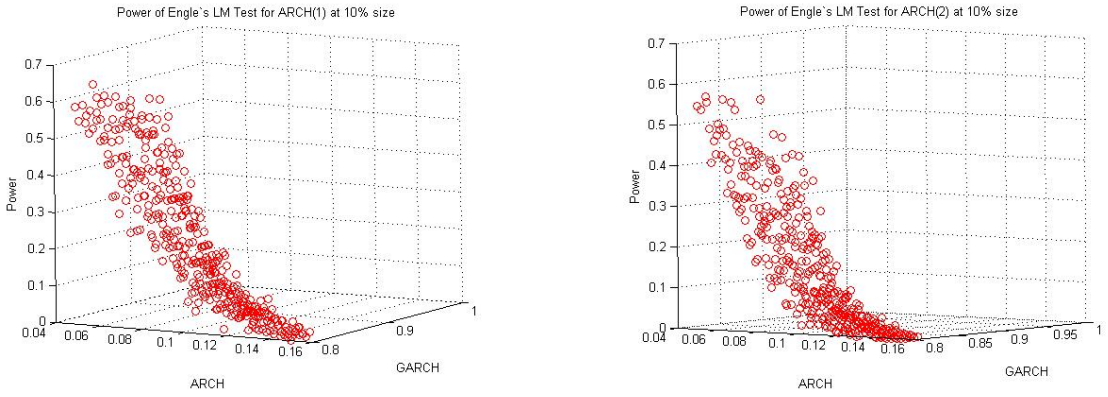


Figure 2.4: Finding a Reasonable Loading Vector

P-values from ARCH LM test on $\widehat{u}_t$, where u_t are errors in the regression $Y_t = \lambda_i f_t + u_t$. The sample size is set to 25,000, $\lambda_i \epsilon \{0, 0.1, ..., 5\}$ and f_t is an ARCH(1) process with $\omega = 0.01$ and $\alpha = 0.50$ in order to approximate the sectors returns used in the empirical application section of this paper.

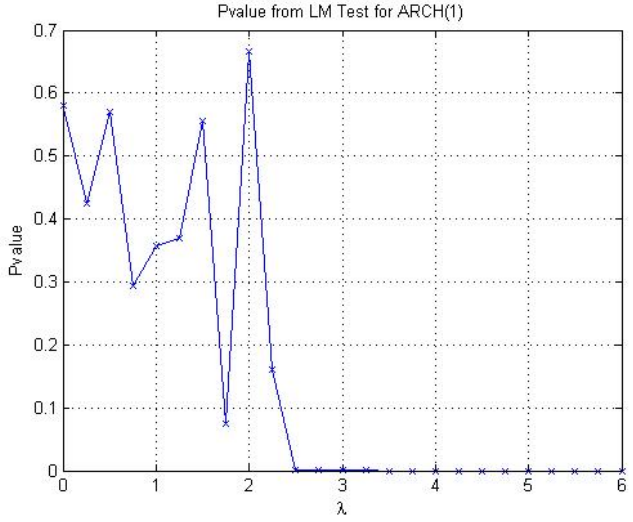


Figure 2.5: Calibrating the Tikhonov Factor

Calibrating the Tikhonov Factor for H_o^3 in a five asset example. The starred line is the normmed % change in the parameter vector from using successively larger Tikhonov Factors during estimation. Choose the Tikhonov Factor such that $\| \frac{\Theta(i+1)}{\Theta(i)} - 1 \| > tol$, where the tolerance (tol) is the solid line in the graph.

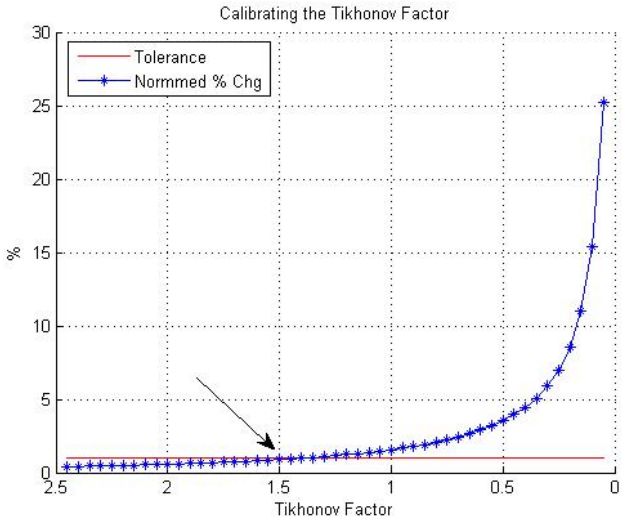


Figure 2.6: Size & Power of Phase 1

Size and Power of over-identified restrictions test in Phase 1 to determine the number of common factors. 250 paths of a five asset system from a single GARCH(1,1) factor with loading vector $\Lambda^a = [10, 9, 8, 7, 6]$ *are simulated. The proportion of sample paths that reject the null of a single common factor is used to compute the empirical size of the test. The power is calculated similarly using a two factor alternative.*

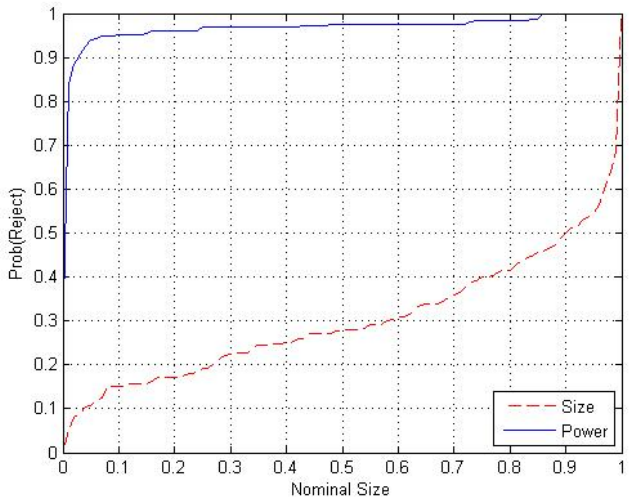


Figure 2.7: Cumulative Sector Returns - Part 1

Cumulative returns of value weighted sector portfolios of the U.S. equity market. Growth of one dollar invested at 01/02/1990 and held through 12/29/2006.

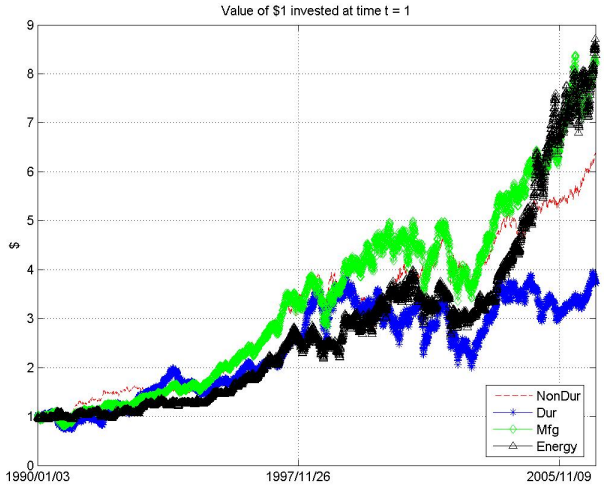


Figure 2.8: Cumulative Sector Returns - Part 2

Cumulative returns of value weighted sector portfolios of the U.S. equity market. Growth of one dollar invested at 01/02/1990 and held through 12/29/2006.

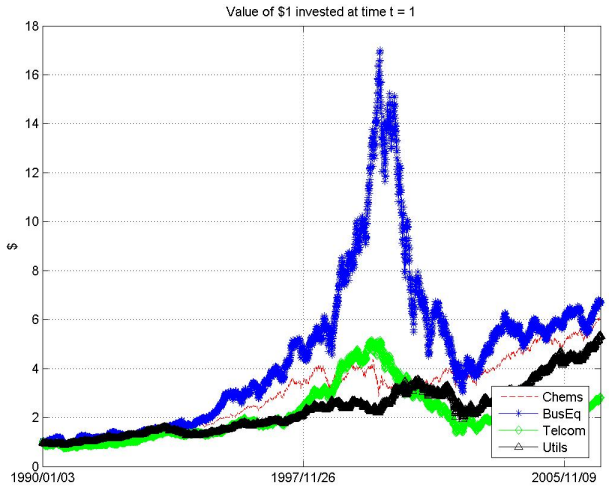


Figure 2.9: Cumulative Sector Returns - Part 3

Cumulative returns of value weighted sector portfolios of the U.S. equity market. Growth of one dollar invested at 01/02/1990 and held through 12/29/2006.

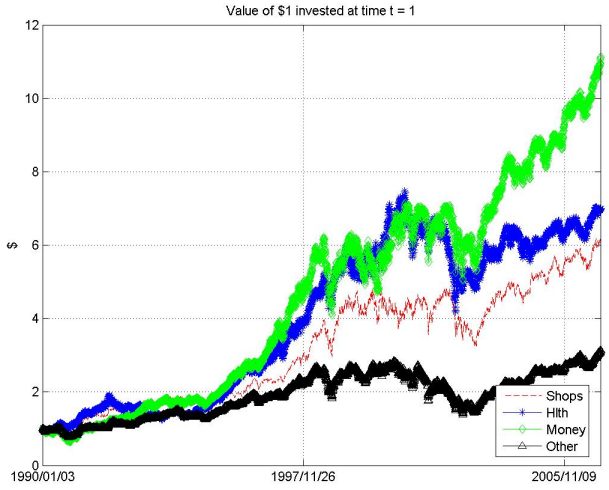


Figure 2.10: # Of Factors Implied by Phase 1

# of factors implied by Phase 1 sequential testing procedure. Computed for a three year estimation period beginning at the dates marked.

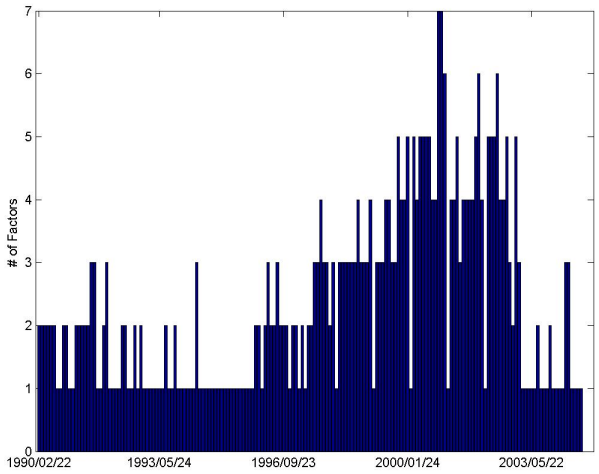


Figure 2.11: Forming Conditionally Homoscedastic Portfolios from 12 Sectors

P-values from ARCH(1) test computed over three period beginning at the dates marked. The solid blue line are the p-values at each date averaged over the 12 sector portfolios. The red stars are the p-values at each date averaged over the $(n-k)$ auxiliary portfolios.

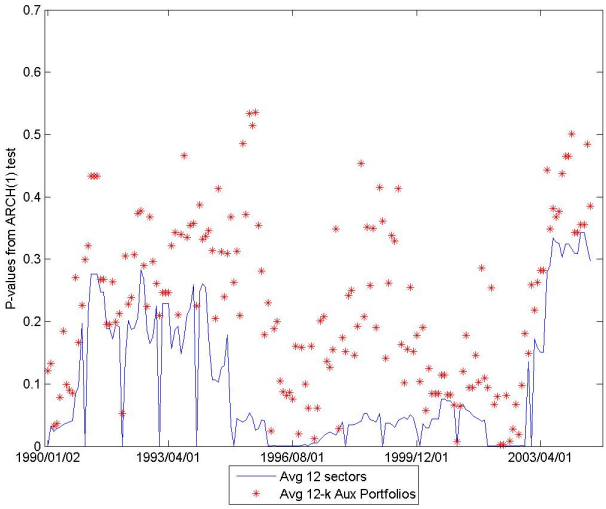


Figure 2.12: Forming Constant Conditional Correlation Portfolios from 12 Sectors

P-values from test for Constant Conditional Correlation computed over three period beginning at the dates marked. The solid blue line are the p-values at each date for the 12 sector portfolios. The red stars are the p-values at each date for the $(n-k)$ auxiliary portfolios.

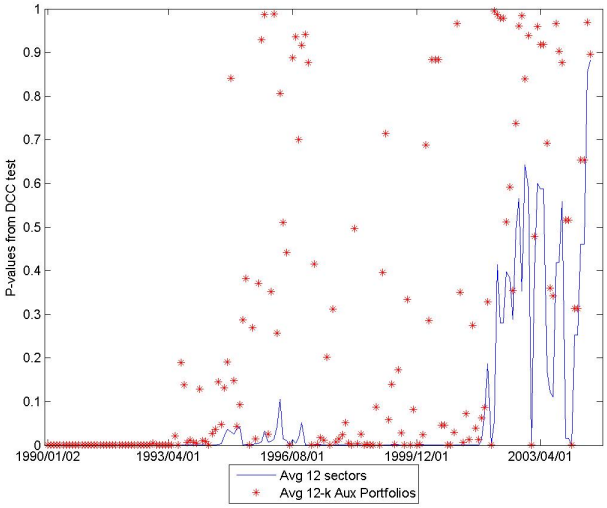


Figure 2.13: Portfolio Performance (Trading Costs = 0.0)

Portfolio Performance for Minimum Volatility investor with Trading Costs = 0.0. Top left panel: average % daily return. Top right panel: average daily standard deviation. Dark green bar: CRSP Value-Weighted market index. Light Green Bar: portfolio implied by DCC model. Yellow Bar: portfolio implied by MVFSV model. Bottom left panel: M2. Bottom right panel: Φ. Dark Green Bar: portfolio implied by DCC model. Yellow Bar: portfolio implied by MVFSV model.

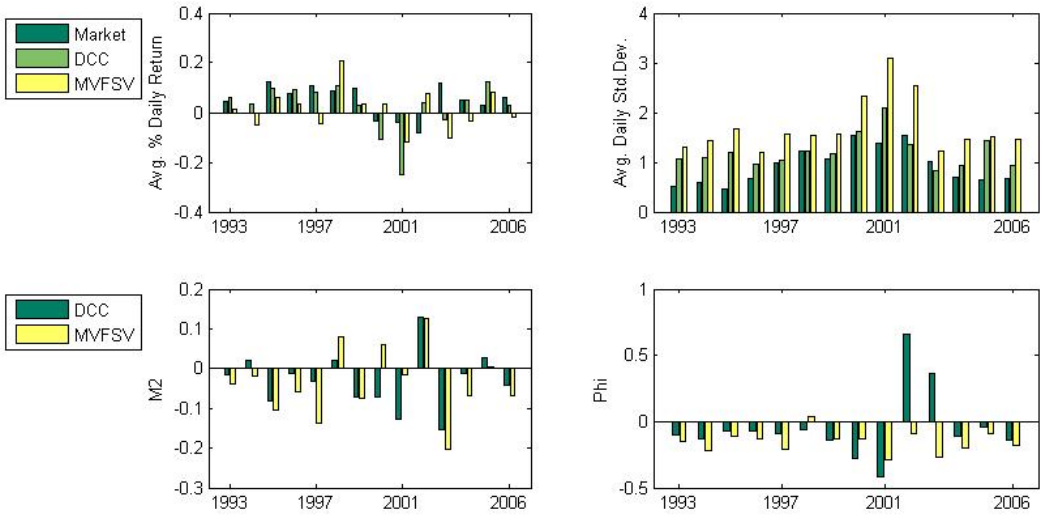


Figure 2.14: Portfolio Performance (Trading Costs = 0.5)

Portfolio Performance for Minimum Volatility investor with Trading Costs = 0.5. Top left panel: average % daily return. Top right panel: average daily standard deviation. Dark green bar: CRSP Value-Weighted market index. Light Green Bar: portfolio implied by DCC model. Yellow Bar: portfolio implied by MVFSV model. Bottom left panel: M2. Bottom right panel: Φ. Dark Green Bar: portfolio implied by DCC model. Yellow Bar: portfolio implied by MVFSV model.

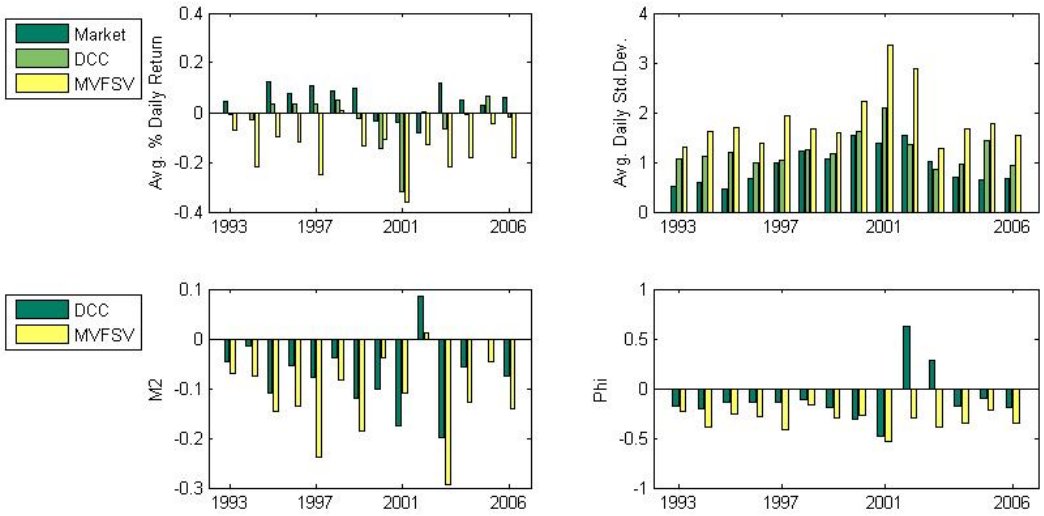


Figure 2.15: Total Portfolio Activity

Total Portfolio Activity: $TPA = \sum_{i=1}^{n} |w_{i,t+1} - w_{i,t}|$, *where* n *is number of assets and* w *are the asset weights.*

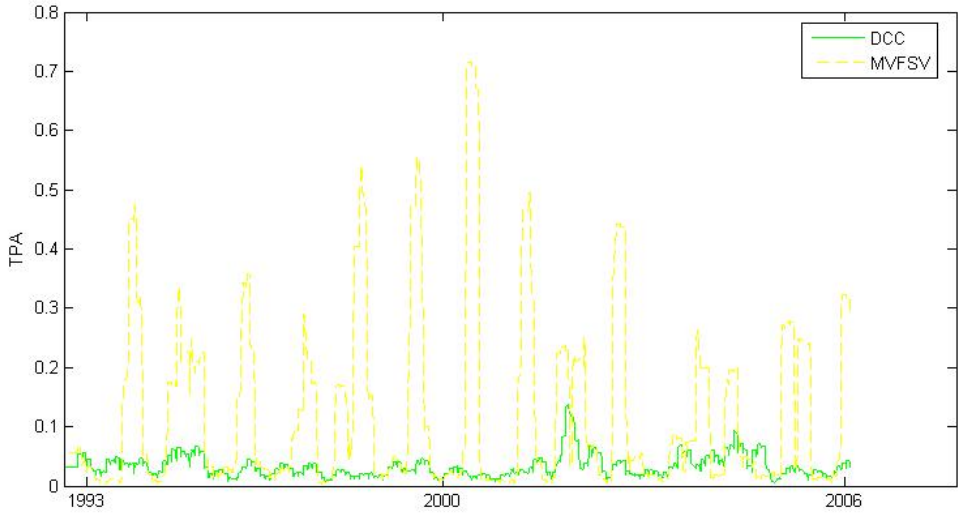


Figure 2.16: Variance Forecasts

Estimation Error for the Variance of the Business Equipment Sector; MVFSV relative to DCC.

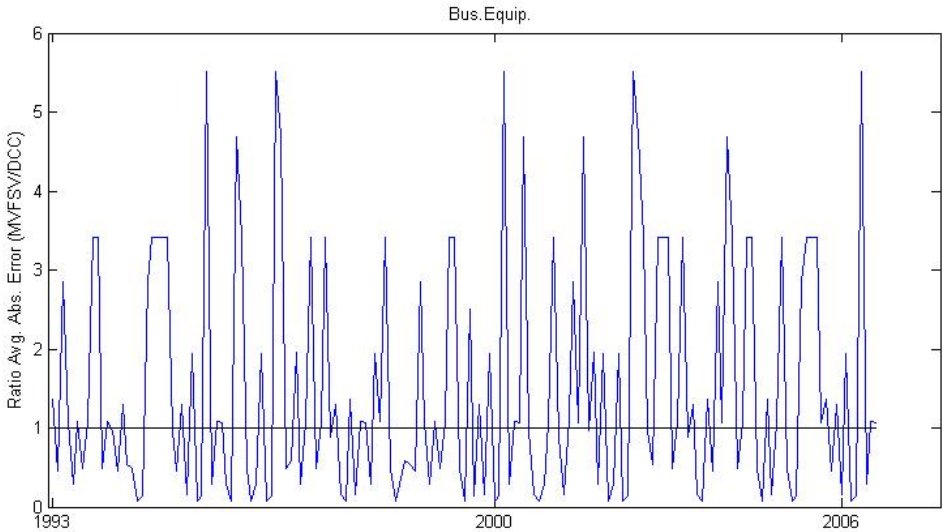
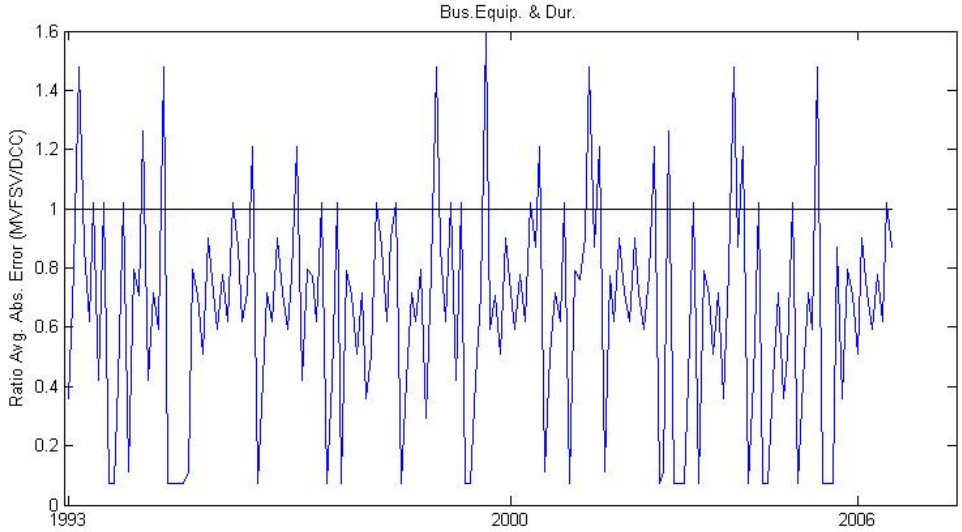


Figure 2.17: Covariance Forecasts

Estimation Error for the Covariance between the Business Equipment and Durable Goods Sectors; MVFSV relative to DCC.



Chapter 3

Indirect Inference for Moment Equations with Infinite Variance

3.1 Introduction

The usual $\sqrt{T}$ consistent and asymptotically normal behavior of GMM estimators relies crucially upon the existence of the variance of the moment conditions. Unfortunately, moment conditions with infinite variance abound the econometrics literature. For example, consider the very simple case of estimating the persistence parameter of an autoregressive process with Student's-T(ν) innovations, where $2 \leq \nu < 4$. Or consider the financial economic example of modeling equity volatility with a GARCH model, where the persistence parameters are outside of the bounds for finite higher order moments as derived by Bollerslev (1986). In either case, GMM estimators may not yield the desired asymptotics. In this paper, we attempt to remedy the issue by employing an Indirect Inference technique in which the moment equations used for estimation are truncated.

The technique begins by truncating the series associated with each moment condition at some predefined threshold. Several truncation techniques are readily available from the field of Robust Statistics, including trimming, Winsorizing, smoothing, and others. This paper focuses on the Winsorizing technique, whereby the "extreme" realizations of the moment conditions are replaced with some pre-defined truncation threshold.

With the Winsorizing technique in place, we then relate the truncated series to the parameters of interest through a binding function that we can estimate via simulation. Since the truncated moment conditions have by definition a finite variance, standard GMM theory applies.

Of course, in sharp contrast with the semi-parametric nature of GMM, the computation of a Monte Carlo assessment of the binding function requires a parametric model to define the simulator. However, as with the Partial Indirect Inference approach of Dridi, Guay and Renault (2007), we are able to check that in many circumstances the estimation of the parameters of interest is not sensitive to the choice of parametric model used in the simulator.

The truncation technique was applied first in a GMM setting in a series of articles by Ronchetti and co-authors[1]. As in our technique, these authors advocate truncating the extreme realizations of the moment conditions. Also similar to our approach, in Genton and Ronchetti (2003), the authors suggest using robust estimators within the simulator of an Indirect Inference approach.

Importantly, however, our technique differs from the Robust GMM approach in its implementation as well as its motivation. With regard to implementation, a key contribution of this work to the literature is the way in which we choose the truncation threshold. Although Ronchetti *et al.* offer some guidelines for this choice, we make explicit the proper truncation threshold by exploiting a tradeoff inherent in various components of the asymptotic variance matrix of the estimator. More importantly, we claim that the Robust GMM approach can not, in general, identify the parameters of interest while still solving the infinite variance problem.

With regard to motivation, the primary focus of the Ronchetti *et al.* work is to robustify GMM estimators and test statistics in the presence of outliers. Viewing the extreme observations in a sample as outside of the true Data Generating Process (DGP) of the data motivates their use of influence functions during the truncation process. In contrast, our approach views the extreme observations as part of the true DGP. As such, we bypass the use of influential statistics and truncate the moment conditions directly.

[1]See Ronchetti and Trojani (2001) for a good overview.

Our application of interest is an AR(1) process whose innovation term has potentially infinite higher order moments. When the DGP of the innovations is symmetric, we find that the truncation procedure is able to restore standard $\sqrt{T}$ Gaussian asymptotics quite easily.

In the face of asymmetric innovations, we find that an additional simulation based tool is needed to correct for the inherent bias of the estimator. Not only are these resulting estimates relatively accurate and approximately Gaussian, but they also appear to be robust to possible mis-specification of the simulator used. In fact, we show that as soon as we are able to mimic reasonably well the skewness of the moment conditions, our estimator performs well, even if the distribution used in the simulator is different from the DGP.

This chapter proceeds as follows. In section 2 we provide a very simple introductory example that describes the problem we face and offers potential solutions. In section 3 we outline the problem of moment equations with infinite variance in a more general setting. The truncation technique is detailed as a solution in section 4. Here, we discuss the various types of truncation possible, detail our Truncated GMM and Indirect Inference estimators, advocate a manner in which to choose the optimal truncation threshold, and discuss the implications of symmetry of the process for bias and parameter identification. Section 5 reports a range of Monte Carlo work illustrating the effectiveness of the truncation technique. Section 6 concludes.

3.2 An Introductory Example

Suppose we have at our disposal a sequence of identically distributed observations

$$y_t = \theta + u_t \tag{3.1}$$

where $u_t, t = 1, ..., T$ have zero mean and follow the probability distribution Q. Our goal is to estimate the location parameter θ.

The true unknown parameter of interest θ^0, is identified as the unique solution of the moment equation

$$E[y_t - \theta] = 0 \tag{3.2}$$

The resulting moment estimator is the sample mean $\bar{\theta} = 1/T \sum_{t=1}^{T} y_t$. Unfortunately, if Q does not have a finite variance, the $\bar{\theta}$ may have a non-standard asymptotic distribution and rate of convergence.

As already announced, we intend to develop an alternative estimator that overcomes the infinite variance problem, without precise knowledge of the distribution Q.

The first step is to eliminate the infinite variance problem by Winsorizing moment condition (3.2). Specifically, for a positive threshold c:

$$\phi_t^{W,c}(\theta) = (y_t - \theta) Min \left[1, \frac{c}{|y_t - \theta|}\right] \tag{3.3}$$

By construction, $|\phi_t^{W,c}(\theta)| \leq c$, and thus the variance of the truncated moment conditions is finite.

Unfortunately, $\phi_t^{W,c}(\theta)$ may not characterize the true unknown parameter of interest, θ^0. More precisely, it is in general possible to define a unique number $\gamma(c)$ as a solution of

$$E\left\{[u_t + \gamma(c)] Min \left[1, \frac{c}{|u_t + \gamma(c)|}\right]\right\} = 0 \tag{3.4}$$

where u_t follows the distribution Q. Note that except if Q is symmetric, there is no reason to believe that $\gamma(c) = 0$. Therefore, truncation may not be a sufficient means by which to achieve our goal of estimating θ. An additional tool may be needed. We sketch three possible approaches:

Method 1: Truncated GMM

The truncated GMM estimator $\hat{\beta}$ is defined as the solution β of the empirical moment equations:

$$(1/T) \sum_{T=1}^{T} (y_t - \beta) Min \left[1, \frac{c}{|y_t - \beta|}\right] = 0 \tag{3.5}$$

Under standard regularity conditions, $\hat{\beta}$ is a consistent estimator of the true unknown value β^0 defined as a solution β of the population moment conditions:

$$E\left\{(y_t - \beta) Min \left[1, \frac{c}{|y_t - \beta|}\right]\right\} = 0 \tag{3.6}$$

Note that the truncated GMM estimator $\hat{\beta}$ is not in general a consistent estimator of the true unknown value of interest θ^0 since, from equation (3.4):

$$\beta^0 = \theta^0 - \gamma(c) \tag{3.7}$$

The Monte Carlo experiments in section 3.5 will assess how the skewness of distribution Q impacts the truncation bias $\gamma(c)$.

Method 2: Robust GMM (Ronchetti and Trojani (2001))

In a more general setting of over-identified GMM, Ronchetti and Trojani (2001) suggests to capture the truncation bias $\gamma(c)$ by introducing an additional degree of freedom γ within the moment equation (3.6). By doing so, they hope to allow this modified moment condition to still yield θ^0 as a solution. Specifically, they propose using the following modified moment condition:

$$E\left\{(y_t - \theta + \gamma)Min\left[1, \frac{c}{|y_t - \theta + \gamma|}\right]\right\} = 0 \tag{3.8}$$

Of course, equation (3.8) will in general accept an infinite number of solutions (θ, γ) such that:

$$\theta - \gamma = \theta^0 - \gamma(c) \tag{3.9}$$

The key hope of Ronchetti and Trojani (2001) is to get the right solution $(\theta^0, \gamma(c))$ as the limit of an algorithm based on equation (3.8) rewritten as:

$$g(\theta, \gamma, \sigma) + \gamma h(\theta, \gamma, \sigma) = 0 \tag{3.10}$$

where

$$g(\theta, \gamma, \sigma) = E\left\{(y_t - \theta)Min\left[1, \frac{c\sigma}{|y_t - \theta + \gamma|}\right]\right\} \tag{3.11}$$

$$h(\theta, \gamma, \sigma) = E\left\{Min\left[1, \frac{c\sigma}{|y_t - \theta + \gamma|}\right]\right\} \tag{3.12}$$

Note that equation (3.10) introduces actually two degrees of freedom within the moment equation (3.6): not only the parameter γ to adjust for the truncation bias, but also a positive

parameter σ to update the truncation threshold. More precisely, Ronchetti and Trojani (2001) proposes the following algorithm:

1. For a given (γ_k, σ_k), θ_k corresponds to the truncated GMM estimator produced from the moment conditions truncated by $c\sigma_k$ and corrected by γ_k as in equation (3.8). Note that in our simple just-identified setting, the population value of θ_k is nothing but the solution of:

$$g(\theta_k, \gamma_k, \sigma_k) + \gamma_k h(\theta_k, \gamma_k, \sigma_k) = 0 \tag{3.13}$$

2. For a given $(\theta_k, \gamma_k, \sigma_k)$, σ_{k+1} is defined by:

$$\sigma_{k+1}^2 = E\left\{(y_t - \theta_k + \gamma_k)^2 Min^2\left[1, \frac{c\sigma_k}{|y_t - \theta_k + \gamma_k|}\right]\right\} \tag{3.14}$$

3. For a given $(\theta_k, \gamma_k, \sigma_k)$, γ_{k+1} is defined as solution of:

$$g(\theta_k, \gamma_k, \sigma_k) + \gamma_{k+1} h(\theta_k, \gamma_k, \sigma_k) = 0 \tag{3.15}$$

It follows obviously from the comparison of (3.13) and (3.15) that the algorithm cannot work in the simple just-identified set up considered here since it does not allow for any updating of γ_k. Even though the over-identified setting of Ronchetti and Trojani (2001) introduces more subtle issues, there is clearly a fundamental flaw due to a degree of indeterminacy similar to equation (3.9). The only way to disentangle θ^0 from $\gamma(c)$ is to bring a piece of additional information, as proposed in the simulation based method considered below.

Note also that the algorithm may actually converge towards the true value if the effective truncation threshold $c\sigma_k$ goes to infinity, perhaps because σ_k goes to infinity as k grows. Indeed, in such a case, the bias $\gamma(c\sigma_k)$ will converge towards zero and the algorithm will fail to solve the infinite variance problem. A detailed study of that possibility is beyond the scope of this paper. However, it does illustrate a key difference between motivations for truncation: the motivation for robustness in front of local contamination (see Ronchetti and Trojani (2001), property 1 page 51), and our motivation to accommodate infinite variance.

Method 3: Simulation Based Truncated GMM

This third method shares some close similarity with the "Robust Indirect Inference" method proposed by Genton and Ronchetti (2003). However, as mentioned above, the motivation for robustness and the motivation for infinite variance are quite different, especially regarding the key issue of choosing a truncation threshold.

The goal of this method is to use simulations in order to alleviate the indeterminacy seen in equation (3.9) and thereby overcome the flaw of Method 2.

Begin by drawing a sequence of i.i.d. random variables $\tilde{u}_t, t = 1, ..., T$ in the probability distribution Q. Equation (3.4) and the law of large numbers permit us to compute $\gamma(c)$ exactly as the unique solution $\gamma = \gamma^*$ of the following equation in terms of its almost sure limit:

$$\lim_{T=\infty} (1/T) \sum_{t=1}^{T} [\tilde{u}_t + \gamma] Min \left[1, \frac{c}{|\tilde{u}_t + \gamma|}\right] = 0 \tag{3.16}$$

With the truncated GMM estimator and approximation of the bias in hand, we can deduce from equation (3.7) a consistent estimate of θ^0 as

$$\hat{\theta} = \hat{\beta} + \gamma^* \tag{3.17}$$

An appealing feature of this approach is that standard GMM theory and asymptotic formula apply. For instance, in this simple case, our estimator $\hat{\theta}$ inherits the $\sqrt{T}$ asymptotic normality and the asymptotic variance of the Truncated GMM estimator. The only role for simulation is to correct for the bias γ^*. Of course, one can build many such estimators by changing the truncation threshold c, which in turn will change both the Truncated GMM estimator $\hat{\beta}$ and the bias correction term γ^*. Our claim is that for the purpose of estimation in the presence of infinite variance, there should be an optimal truncation threshold, c^*, that minimizes the asymptotic variance. The argument supporting this claims is as follows.

On the one hand, increasing c should decrease the asymptotic variance of the truncated GMM estimator $\hat{\beta}$, and hence $\hat{\theta}$. This is because, for a given long term variance, the moment

conditions in equation (3.6) become more informative about β as c increases. The improved informational content is reflected as an increase in the absolute value of the derivative of the moment conditions with respect to β.

On the other hand, due to the infinite variance problem, increasing c too much will lead us to incorporate the tails of the moment conditions, which is the area where the variance becomes explosive. Beyond some optimal value of the truncation threshold c^*, the moment condition's gain in informational content about β will be overwhelmed by the increase in its variance.

Therefore, as confirmed by our Monte Carlo experiments, we expect our estimator to improve as c increases from 0, and then to deteriorate beyond some optimal value c^*.

Several additional remarks regarding the practical relevance of our method are in order:

1. The application of our method requires simulating in the probability distribution Q. This may be seen as a shame since one important advantage of GMM is precisely the semi-parametric feature of not requiring a full specification of the probability distribution that defines the DGP. This actually is a general weakness of the Indirect Inference method as devised by Gourieroux, Monfort and Renault (1993). Both the simulation based truncated GMM method developed here and the Robust Indirect Inference method of Genton and Ronchetti (2003) are subject to this critique. Even though Dridi, Guay and Renault (2007) provide some general tools to make Indirect Inference slightly more semi-parametric, this potential weakness must be carefully studied on a case by case basis.

2. As far as asymptotic distributional theory is concerned, all estimation methods fail to be fully semi-parametric in the presence of infinite variance. Standard methods often lead to non-standard rates of convergence and asymptotic distributions that depend highly on the probability distribution Q. Our method, on the other hand, generates a dependence on Q through the computation of the bias $\gamma(c)$. It is then an empirical question to decide which of these dependencies is the most detrimental to the application at hand. Monte Carlo experiments suggest that, when properly devised, our method is rather robust

to specification errors in the probability distribution Q. We illustrate that the bias is indeed zero when the DGP is symmetric. In addition, we provide some evidence that any simulator Q, even different from the DGP, will perform properly as soon as we are able to mimic reasonably well the skewness of the truncated moment conditions under the true DGP. Moreover, if an alternative consistent estimator with a non-standard rate of convergence is available, an Hausman-type test may allow us to check formally that our $\sqrt{T}$ consistent estimator actually converges toward the true unknown value θ^0[2].

3. The simulation based method for bias correction can be applied properly with only a finite number of simulations. For instance, let us draw a simulated path $\{\tilde{u}_t\}$ in the probability distribution Q. Assume that, perhaps due to computational constraints, we want to limit the length of the simulated path so that it is only H times longer than the observed path $y_t, t = 1, ..., T$. We have then $\tilde{u}_t, t = 1, ..., TH$. Consider $\hat{\gamma}_{TH}$ defined as the solution of the simulated moment equation:

$$(1/TH)\sum_{t=1}^{TH}[\tilde{u}_t + \gamma]Min\left[1, \frac{c}{|\tilde{u}_t + \gamma|}\right] = 0 \tag{3.18}$$

Regardless of the chosen value for the positive integer H, as T goes to infinity, γ_{TH} will be a consistent estimator of $\gamma(c)$. Therefore, we get a $\sqrt{T}$ consistent and asymptotically normal estimator of θ^0 by computing

$$\hat{\theta} = \hat{\beta} + \hat{\gamma}_{TH} \tag{3.19}$$

4. The finite number of simulations H plays only a minor role in the asymptotic variance of our estimator. Let us denote by V/TH the asymptotic variance of $\hat{\gamma}_{TH}$. It is then clear that the asymptotic variance of $\hat{\beta}$ is V/T, so that the asymptotic variance of our estimator $\hat{\theta}$ is simply $(1/T)(1 + 1/H)V$. Once we move beyond say $H = 5$, an increase

[2]In this regard, we have in mind the truncation based estimator designed by Hill and Renault (2008). In future work we intend to employ a Hausman test comparing our estimator to theirs. See Appendix B for a brief overview of our intended strategy.

in the number of simulations does not alter substantially the asymptotic variance of $\hat{\theta}$. However, our Monte Carlo experiments reveal that a sufficiently large H may be more important for the finite sample properties of $\hat{\theta}$.

Note also that if the distribution used in the simulator Q differs from that of the DGP, then we should alter the computation of the variance matrix. As per Dridi, Guay and Renault (2007), the variance of $\hat{\theta}$ must rather be computed as $(1/T)[V(DGP)+V(Q)/H]$, where $V(DGP)$ is the variance of $\hat{\theta}$ under the DGP, and $V(Q)$ is the variance of $\hat{\theta}$ under the simulated distribution Q.

The general Indirect Inference method derived in this paper for dealing with moment conditions of infinite variance corresponds to simple cases, such as that described in this introductory example. However, when we consider a more general context, we will have to devise an extension of the above approach to accommodate for the fact that the truncation bias depends in general on the true unknown value θ^0 of θ. This may force us to use iteratively the simulation based moment conditions device along a sequence of values of θ.

3.3 Outlining the Infinite Variance Problem

We are interested in estimating a vector of unknown parameters $\theta \in \Theta \subset \Re^q$, from a set of p moment conditions

$$E\left[\phi_t(\theta)\right] = 0 \tag{3.20}$$

These moment conditions are assumed to identify the true unknown value θ^0 of θ:

$$E\left[\phi_t(\theta)\right] = 0 \Leftrightarrow \theta = \theta^0 \tag{3.21}$$

For simplicity, we consider only just-identified moment conditions; i.e. $p = q$. Moreover, the moment conditions generally can be considered as $\phi_t(\theta) = \phi(y_t, y_{t-1}, ..., y_{t-l}, \theta)$, where y_t is a stationary and ergodic process.

Let us assume for the moment that all the regularity conditions detailed in Hansen (1982) are fulfilled. Then, for any sequence (W_T) of positive matrices whose limit (W) is positive

definite, a consistent GMM estimator $\hat{\theta}_{GMM}$ is computed as

$$\hat{\theta}_{GMM} = \underset{\theta}{\operatorname{argmin}} \ \bar{\phi}_T(\theta)' W_T \bar{\phi}_T(\theta) \tag{3.22}$$

where $\bar{\phi}_T(\theta) = 1/T \sum_{t=1}^{T} \phi_t(\theta)$.

The resulting GMM estimator is asymptotically normal with a $T^{1/2}$ rate of convergence and a variance matrix that does not depend on W. In particular, $T^{1/2}(\hat{\theta}_{GMM} - \theta^0) \xrightarrow{d} N(0, [M'\Omega^{-1}M]^{-1})$ where $M = E\left[\frac{\partial \phi_t(\theta)}{\partial \theta}'\right]$, and Ω is the long term covariance matrix of the moment conditions.

As already announced, the purpose of this paper is to accommodate standard GMM theory when all but possibly one of the regularity conditions outlined in Hansen (1982) are fulfilled. Specifically, we may find ourselves in a situation where all components of $\phi_t(\theta^0)$ are integrable, yet some may not be square integrable. Note, however, that the fact that some terms of Ω become explosive does not necessarily imply an infinite variance for $T^{1/2}(\hat{\theta}_{GMM} - \theta^0)$. In fact, it may be the case that $T^{1/2}(\hat{\theta}_{GMM} - \theta^0)$ converges almost surely to zero.

Consider the simple AR(1) process: $y_t = \omega + \rho y_{t-1} + u_t$, which is the application of interest in this paper. We assume $|\rho| < 1$ and that the innovation term u_t comes from an i.i.d. sequence of symmetrically distributed random variables whose distribution function F satisfies:

$$\lim_{t=\infty} \ t^{\alpha}[1 - F(t)] = k \tag{3.23}$$

for some $k > 0$, and $\alpha \in (0, 2)$. These u_t are said to be symmetric and asymptotically Pareto of index α. The symmetric stable distribution of index α is a particular case. This structure suggests that $|u_t|^{\gamma}$ is integrable only for $\gamma < \alpha$, which implies that it has infinite variance. Kanter and Steiger (1974) show that, for the case of $\omega = 0$, the only stationary solution to y_t is well defined as the almost sure limit of its moving average representation. That solution, also has infinite moments of order $\geq \alpha$.

Within this framework Hannan and Kanter (1977) are able to show that for any $\gamma > \alpha$, $T^{1/\gamma}(\hat{\rho}_{GMM} - \rho^0)$, and in particular $T^{1/2}(\hat{\rho}_{GMM} - \rho^0)$, converges almost surely to zero.

The intuition for this result is described nicely by Cox (1966). Take an isolated very large value u_t, which is more likely if α is small. The presence of this extreme observation will result in a relatively large value of $y_t = \rho y_{t-1} + u_t \approx u_t$, and of $y_{t+1} = \rho y_t + u_{t+1} \approx \rho u_t$. Through their ratio, the two consecutive observations (y_t, y_{t+1}) characterize the value of ρ almost exactly. Moreover, this result can be seen from the perspective that the extreme observation is influential, and thus can be quite informative about the true unknown value $\phi_t(\theta^0)$. However, it would actually take a great deal of trust in the model specification to accept that the estimator basically is produced by a couple of extreme observations. We will on the contrary, reduce / eliminate the impact of such observations in order to be sure that the full sample is taken into account into the estimation process.

Regardless of the intuition, we know that fatter tails are associated with smaller values for α and faster rates of convergence of the GMM estimator. This situation is not user friendly. Even before we face the difficulty of identifying the limit distribution of the estimator, characterizing the rate of convergence requires knowledge of the exact value of α, which is at odds with the semi-parametric nature of GMM.

Note that in the above discussion, we never question the integrability of the moment conditions. Thus, while standard asymptotic theory for OLS is invalid, the theory of Least Absolute Deviations (LAD) may work.

More generally, M-Estimators, such as least squares and LAD, have been explored in the extreme setting of infinite variance, and the even more extreme setting of infinite mean, by many authors. For instance, see Davis, Knight and Liu (1992), and the references therein, for a discussion on estimating the parameters of autoregressive models where standard asymptotics do not apply. The authors consider an autoregressive process with innovations residing in the Domain of Attraction of a Stable Law, where the tail index is less than two[3]. They are able

[3]To be a bit more precise, the authors consider

$$P[|u_t| > x] = x^{-\alpha} L(x)$$
$$lim_{x\to\infty} \frac{P[u_t > x]}{P[|u_t| > x]} = p, \ 0 \leq p \leq 1$$

where $L(x)$ is a slowly varying function. Note: A positive, Lebesgue-measurable function L on $(0, \infty)$ is slowly varying at ∞ if $lim_{x\to\infty} \ L(tx)/L(x) = 1, t > 0$. See McNeil, Frey and Embrechts (2005) page 268.

to overcome the problem of applying a standard Central Limit Theorem by finding a rate of convergence $a_T = T^{1/\alpha}L_1(y)$, where $L_1(y)$ is some slowly varying function. Importantly, as the authors note, the limiting distribution of their estimators is difficult to compute and highly dependent upon the distribution of the innovations. Moreover, despite the fact that a_T may offer a faster rate of convergence than $\sqrt{T}$, this rate may be unknown, making such an approach impractical.

Quasi maximum likelihood estimators (QMLE) in the face of infinite unconditional variance have also been explored. Authors such as Ling and McAleer (2003) and Francq and Zakoian (2004) explored the asymptotics of ARMA-GARCH models with infinite variance. Francq and Zakoian (2004) are able to establish the consistency and asymptotically normality of the QMLE as long as they maintain some minor restrictions on the existence of moments of the innovation process. Ling (2007) and Lange, Rahbek and Jensen (2007) also explore the infinite variance problem in the ARMA-GARCH setting. However, these authors attempt to relax the restrictions in Francq and Zakoian (2004) by truncating the likelihood functions.

This paper is motivated by similar issues and incorporates truncation, akin to Ling's Self Weighted QMLE and Lange, Rahbek, and Jensen's Modified QMLE. In fact, this chapter of my dissertation is intended to establish the framework upon which we can explore the infinite variance problem with an ARMA-GARCH setting while using standard GMM theory. Specifically, we propose to truncate the moment conditions in a manner to reduce/eliminate the impact of extreme observations, thereby restoring Gaussian asymptotics and the standard rates of convergence.

3.4 A Truncation Solution

We propose overcoming the estimation problem detailed in the previous section by truncating the moment conditions in such a way so as to yield a finite variance.

3.4.1 Types of Truncation

Truncation, in one form or another, is well established in the field of Robust Statistics. Drawing inspiration from this line of literature, we consider several forms of truncation. First, we could trim away those realizations of the moment conditions whose values lie beyond positive truncation threshold. Let $c = c(i)$ for $i = 1, ..., p$, be a $p-$dimensional vector of truncation thresholds whose components all are positive. We define the "trimmed" moment conditions for each $i = 1, ..., p$ as $\phi_{i,t}^{tr,c(i)}(\theta) = \phi_{i,t}(\theta) I_{|\phi_{i,t}(\theta)|<c(i)}$[4]. Moreover, we can form the vector of trimmed moment conditions as $\phi_t^{tr,c}(\theta) = [\phi_{i,t}^{tr,c(i)}(\theta)]_{1\leq i\leq p}$. The sample mean of the trimmed moment conditions, which is key for GMM estimation, is $\bar{\phi}_T^{tr,c}(\theta) = 1/T \sum_{t=1}^{T} \phi_t^{tr,c}(\theta)$.

In a second truncation technique, we could, generally speaking, replace the extreme observations with the chosen truncation threshold. We define the Winsorized version of moment condition (i) as $\phi_{i,t}^{W,c(i)}(\theta) = \phi_{i,t} Min\left[1, \frac{c(i)}{|\phi_{i,t}(\theta)|}\right]$. We can form the vector of Winsorized moment conditions as $\phi_t^{W,c} = [\phi_{i,t}^{W,c(i)}]_{1\leq i\leq p}$. Furthermore, by a slight abuse of notation, we can choose not to truncate moment condition (i) by setting $c(i) = +\infty$. The sample mean of the Winsorized moment conditions may be written as $\bar{\phi}_T^{W,c}(\theta) = 1/T \sum_{t=1}^{T} \phi_t^{W,c}(\theta)$.

Notice that these truncation techniques introduce a degree of non-smoothness to the moment conditions. As a result, another key regularity condition typically required for GMM estimation and inference may be violated. See for instance, Hall (2005) for an excellent review of these conditions. Specifically, the derivative matrix $\partial \phi_t^{W,c}(\theta)/\partial\theta$ may not exist for all points in the sample space.

There are at least two possibilities for proceeding with our moment-based estimation. First, we could use a smoothed truncation technique. For instance, Vetzal (1992) overcomes the discontinuities by approximating the moment conditions in small neighborhoods of the discontinuity points, in our case the truncation threshold (c), with a sixth order polynomial.

[4]We use this notation loosely. In fact, the definition of our trimmed series is as follows. Suppose $\phi_{i,t}$ for $t = 1, ..., T$ belongs to a distribution F, which is strictly increasing and symmetric around 0. Consider the order statistics $X_1 < X_2 < ... < X_T$ of $\phi_{i,1}, \phi_{i,2}, ..., \phi_{i,T}$. Consider some $c(i) > 0$ and define $F(c(i)) = \alpha$. Modifying slightly the definitions in Bickel (1965), define the c(i)-trimmed series associated with $\phi_{i,t}$ as $\phi^{tr,c(i)} = \left[X_{[\alpha T]+1}, ..., X_{T-[\alpha T]}\right]$, where $[\alpha T]$ is the greatest integer αT for $0 \leq \alpha < 1/2$. The $c(i)$-trimmed mean can then be defined as $\bar{\phi}_{i,T}^{tr,c(i)} = [T - 2[\alpha T]]^{-1} \sum_{i=[\alpha T]+1}^{T-[\alpha T]} X_i$. This notation helps to ensure that we are dropping the offending observations, and not simply replacing them with zero values.

The derivative is then based on this smoothed polynomial approximation of the moment condition. This would ensure that we use differentiable moment conditions with finite variance, so that standard GMM theory would apply. However, we leave this for future work.

Instead, we rely upon the empirical process work of Pakes and Pollard (1989), Andrews (1994) and McFadden and Newey (1994) in order to use the Winsorized moment conditions directly. These authors suggest that although non-smoothness of the sample moment conditions could be a hindrance, their expectation may be sufficiently smooth so as to yield Gaussian asymptotics at standard rates of convergence.

Re-write the Winsorized moment conditions as a sum of indicator functions as follows: $\phi_{i,t}^{W,c(i)}(\theta) = -c(i) \times 1_{\{\phi_{i,t}(\theta) \leq -c(i)\}} + \phi_{i,t}(\theta) 1_{\{-c(i) < \phi_{i,t}(\theta) < c(i)\}} + c(i) \times 1_{\{\phi_{i,t}(\theta) > c(i)\}}$. From this perspective, it is easy to see that the points of discontinuity reside exclusively with the extreme observations, which are rare by design of the problem. In fact, continuity of the probability distribution of y_t is sufficient to ensure that the truncated moments do not place any mass at the points $\pm(c)$. Therefore, the set of points where $E[\partial \phi_t^{W,c}(\theta)/\partial\theta]$ doesn't exist has Lebesgue measure zero. We can proceed unhindered[5].

3.4.2 Truncated GMM and Indirect Inference

Consider a very general time series process of the form:

$$y_t = g(y_{t-1}, u_t, \theta) \tag{3.24}$$

$$u_t = r(u_{t-1}, \varepsilon_t, \theta) \tag{3.25}$$

where $g(\cdot)$ is a continuous and at least twice differentiable function. The process $\{\varepsilon_t\}$ is i.i.d. with known distribution Q and the process y_t is stationary, with $\theta \in \Theta \subset \Re^q$. Following Gourieroux, Monfort and Renault (1993), this setup can be extended to include exogenous variables and multiple lags of y_t and u_t. The importance of equations (3.24) and (3.25) is that they allow the process y_t to be simulable. That is, we can draw simulated values $\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_T$

[5] Hall (2005) describes the work of Melino and Turnbull (1990) as precedent for this approach.

according to Q. Given initial values y_0 and u_0, we can then build

$$\begin{aligned}
\tilde{y}_0(\theta, y_0, u_0) &= y_0 \\
\tilde{u}_t(\theta, u_0) &= r\left(\tilde{u}_{t-1}(\theta, u_0), \tilde{\varepsilon}_t, \theta\right) \\
\tilde{y}_t(\theta, y_0, u_0) &= g\left(\tilde{y}_{t-1}(\theta, y_0, u_0), \tilde{u}_t(\theta, u_0), \theta\right)
\end{aligned}$$

from which we can compute the simulated values

$$\tilde{y}_t(\theta, y_0, u_0), \ t = 0, ..., T$$

This general framework allows us to consider simple linear and autoregressive models, or more complex structures such as GARCH models.

Consider again the vector of truncated population moment conditions $E\left[\phi_t^{W,c}(\theta^0)\right] = 0$, where $\phi_t(\theta) = \phi(y_t, ..., y_{t-l}, \theta)$, for some fixed function ϕ. As we will discuss in section 3.4.4, the truncation device may generate an asymptotic bias, implying that the truncated population moment condition will not hold, and this model will be unable to identify the parameter of interest (θ).

We claim that the truncated population moment condition instead will identify uniquely an auxiliary parameter β. In other words, $E\left[\phi^{W,c}(\beta^0)\right] = 0$, where β^0 may not necessarily be equal to θ^0. We assume that the $dim(\beta) = dim(\theta)$. An extension of this procedure to accommodate over-identified moment conditions about θ is beyond the scope of this paper. Moreover, we assume that there is a natural one to one correspondence between the elements of β and those of θ. The relation between β and θ can then be seen as $\beta = b(\theta)$, where $b(\cdot)$ is referred to as the binding function. If $b(\cdot)$ was known, we could estimate β, and recover θ simply by inverting that function. Unfortunately, the binding function is unknown. However, thanks to the Indirect Inference approach of Gourieroux, Monfort and Renault (1993), we can estimate this binding function via simulations and recover θ. As is standard in Indirect Inference, there are three critical steps.

In step one, we estimate the auxiliary parameter β with the truncated moment conditions

upon the observed data. We refer to this approach simply at Truncated GMM. Specifically, we solve:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \ ||\bar{\phi}_T^{W,c}(\beta)|| \tag{3.26}$$

Since our empirical applications will remain in a just-identified setting, the norm in the optimization problem above can be a simple quadratic without an explicit weighting matrix. As outlined in previous sections, these moment conditions are differentiable in expectation, and have finite variance thanks to truncation. Moreover, the balance of the regularity conditions outlined by Hansen (1982) are satisfied. Therefore, standard GMM asymptotics apply. We have $T^{1/2}(\hat{\beta} - \beta^0) \xrightarrow{d} N(0, Var(\hat{\beta}))$, where $Var(\hat{\beta}) = \left[\frac{\partial \phi_t^{W,c}(\beta)'}{\partial \beta} Var(\phi_t^{W,c}(\beta))^{-1} \frac{\partial \phi_t^{W,c}(\beta)}{\partial \beta'}\right]^{-1}$.

In step two, we estimate β via Truncated GMM on simulated paths of y. The process $\tilde{y}_t, t = 1, ..., T$ is simulated according to the DGP outlined in equations (3.24) and (3.25). The simulation requires knowledge of the distribution of the innovations term ε_t. However, one of the virtues of Indirect Inference is that it allows us to be robust to mis-specifications of this innovation term. We will provide Monte Carlo evidence in support of this claim in section 3.5.

The auxiliary parameter is then estimated via Truncated GMM on a single path of the simulated process:

$$\tilde{\beta}_h(\tilde{y}_h, \theta) = \underset{\beta}{\text{argmin}} \ ||\bar{\phi}_T^{W,c}(\beta(\tilde{y}_h, \theta))|| \tag{3.27}$$

These estimates are averaged over H simulated paths; $\tilde{\beta}(\tilde{y}, \theta) = 1/H \sum_{h=1}^{H} \tilde{\beta}_h(\tilde{y}_h, \theta)$.

In step three, we match the two estimates of the auxiliary parameter in order to recover the parameter of interest. The resulting Indirect Inference estimator is computed as

$$\hat{\theta}_{II} = \underset{\theta}{\text{argmin}} \ ||\hat{\beta} - \tilde{\beta}(\tilde{y}, \theta)|| \tag{3.28}$$

Following Gourieroux, Monfort and Renault (1993), this estimator follows standard $\sqrt{T}$ Gaussian asymptotics.

$$\sqrt{T}(\hat{\theta}_{II} - \theta^0) \xrightarrow{d} N(0, \Sigma_{II}) \tag{3.29}$$

$$\Sigma_{II} = \left(1 + \frac{1}{H}\right) \left[\frac{\partial b'}{\partial \theta} Var(\hat{\beta})^{-1} \frac{\partial b}{\partial \theta'}\right]^{-1} \tag{3.30}$$

Notice that the Truncated GMM and Indirect Inference estimators assume prior knowledge of the vector of truncation thresholds (c). In the next section we explore carefully the proper choice of threshold.

3.4.3 Choice of Truncation Threshold

The threshold at which to truncate has heretofore been somewhat of an unresolved issue in the literature. For instance, Ling (2007) offers no guidance for the choice of truncation threshold in his Self-Weighted QMLE. Lange, Rahbek and Jensen (2007) are able to offer some guidelines for their Modified QMLE. The authors suggest trimming at most 5% of the terms in the likelihood function. However, as the authors note, this guideline is based solely on simulation evidence and is calibrated specifically for the AR-ARCH model. Moreover, their threshold choice may depend on the parameter of interest to the researcher.

The Robust GMM work of Ronchetti and Trojani (2001) addresses the choice of truncation threshold in a rather subjective manner. That is, they leave the choice of threshold open to the econometrician. Specifically, they advocate choosing a truncation threshold that balances the tradeoff between contamination of outliers expected by the researcher in a given sample and the maximal bias the researcher is willing to accept.

In this paper, we advocate choosing the truncation threshold that minimizes the asymptotic variance of the estimator.

Consider a very simple case where the moment condition is one dimensional, there is a single parameter of interest, and the truncation threshold (c) is a scalar. Assuming for a moment that we have an infinite number of simulation, $H = \infty$, the asymptotic variance of θ in equation (3.30) can be re-written simply as $\Sigma_{II} = Var(\hat{\beta}) \left[\partial b / \partial \theta\right]^{-2}$.

Note further that $\hat{\beta}$ is itself a GMM estimator, whose variance also consists of two pieces: $E[\partial \phi_t^{W,c} / \partial \beta]$, and $Var[\phi_t^{W,c}]$. The variance of θ then can be seen as the amalgamation of three components, written as $\Sigma_{II} = Var[\phi_t^{W,c}] \times \left[\frac{\partial \phi^{W,c}}{\partial \beta}\right]^{-2} \times \left[\frac{\partial b}{\partial \theta}\right]^{-2}$. The relative magnitude of these components dictates how Σ_{II} behaves as (c) varies.

Consider a given level of $Var[\phi_t^{W,c}]$. Increasing (c) from a small value will have two effects that combine to lower the variance of θ. First, an increase in (c) improves the informational

content of the truncated moment conditions regarding the parameter β. This is reflected as an increase in the absolute value of the derivative of the truncated moment conditions with respect to β, which lowers Σ_{II}.

Second, the slope of the binding function $\partial b/\partial\theta$ also will increase with (c), thereby causing Σ_{II} to lower as well. This can be understood best by recognizing that within an Indirect Inference framework, the binding function relates the parameters of interest to the auxiliary model. The slope of the binding function gauges the informativeness of that relationship. A slope of zero indicates that the binding function is useless at recovering the parameter of interest. By contrast, a large slope tells us that the relationship is quite informative. For this reason, we hope to maximize the slope of the binding function.

If $c = 0$, the truncated series effectively would contain no observations, and thus, the auxiliary model would be unable to provide any information with which to recover the parameter of interest. As the truncation threshold increases, for a given value of θ, the auxiliary model becomes more informative and the absolute value of the slope of the binding function increases. This will continue as long as the truncation is effective. Beyond some threshold, for instance $c = Max[\phi_t]$, truncation is irrelevant, and thus the binding function will have achieved maximal slope for that given value of θ.

There is, however, a counterbalancing force. As (c) increases $Var[\phi_t^{W,c}]$ becomes explosive. Beyond some optimal truncation threshold, say c^*, the increase in variance dominates the relationship.

The tradeoff is clear. As c increases, $\partial b/\partial\theta$ and $\partial\phi_t^{W,c}/\partial\beta$ combine to lower Σ_{II}. Yet, as c increases, $Var[\phi_t^{W,c}]$ raises Σ_{II}. We exploit this tradeoff by choosing the value of c that minimizes the sample estimate of Σ_{II}. In section 3.5 we will use Monte Carlo analysis to illustrate the effectiveness of this choice.

In a multivariate setting, the optimal threshold again is chosen by balancing the "size" of the variance of the truncated GMM estimator and the slope of the binding function. Of course, we are in a multivariate setting, so "size" is measured via norms of the relevant components. The logic of the tradeoff as the truncation threshold rises then carries through nicely.

3.4.4 Identification, Bias, and Symmetry

Proper identification of the parameters of interest depends in part upon the symmetry of the truncation technique as well as that of the innovation term. The Winsorizing technique is symmetric by design. It replaces large and small "extreme" observations in equal proportion. Thanks to the intuitive results of Bickel (1965) and Stigler (1973), we know that the Winsorized sample mean is a consistent and asymptotically normal estimator of the population mean on symmetrically distributed data. Therefore, if the process y has symmetric innovations, the truncation technique will not generate an asymptotic bias. The auxiliary parameter β will coincide with the parameter of interest θ. The binding function will be a trivial identity function. There will be no need for an Indirect Inference approach. Estimation and inference can be conducted with a simple Truncated GMM strategy using the truncated population moment conditions $E\left[\phi_t^{W,c}(\theta^0)\right] = 0$.

However, when the innovation terms are asymmetrically distributed, the Winsorizing technique likely will yield an asymptotic bias. For example, consider a simple one dimensional moment condition that is heavily skewed. The population expectation is not a proper measure of the central tendency of the data, and it will be unable to provide an unbiased estimator of the parameter of interest. Consider now a truncation threshold that is used to Winsorize that moment condition at values c and $-c$. The Winsorized moment condition will do little eliminate to the bias. Consequently, we can no longer rely upon the consistency results of Bickel (1965) and Stigler (1973)[6]. Indirect Inference can be used to adjust for this asymptotic bias.

Interestingly, the Indirect Inference approach may have some merit beyond this asymptotic argument. Monte Carlo work in section 3.5 suggests that estimates generated through an Indirect Inference approach can offer a finite sample advantage over the Truncated GMM estimators, by producing less finite sample bias and smaller levels of standard deviation.

[6]Of course, there may be some advantage to invoking an asymmetric truncation scheme in the face of asymmetric innovations. However, we leave this for future work.

3.5 Monte Carlo Evidence

In this section we present a range of Monte Carlo evidence to illustrate the effectiveness of the moment-based truncation technique at overcoming the infinite variance problem described in the previous sections.

The model of interest is a first-order autoregressive process:

$$y_t = \omega + \rho y_{t-1} + u_t \tag{3.31}$$

where u_t is zero mean and i.i.d. according to a Student's T distribution with tail index ν, which we write as $u_t \sim StT(\nu)$.

Ascribing a Student's T distribution for the innovation term is quite useful for our purposes. First, its symmetry implies that the innovations, and thus the moment conditions, also are symmetric. Therefore, as described in section 3.4.4, the Winsorization technique will not generate an asymptotic bias. We also will explore the impact of asymmetric innovations towards the end of the following two sub-sections

A second appealing feature of the Student's T distribution is that the innovations, and thus some multiple of the moment conditions, will have finite moments only up to the degree of the tail index. Specifically, $E[u_t^\gamma] < \infty$ for any $0 < \gamma < \nu$. So, if $\nu = 4$, only the first three moments of u_t are finite. If $\nu = 3$, only the first two moments are finite. And so on. This property lends itself nicely to the design of the Monte Carlo experiments since the existence of the moments of the innovation term is at the heart of our estimation problem.

In addition, the Student's T distribution characterizes well a practical dilemma. Researchers often are fond of the Student's T because its heavy tails capture well the behavior of series encountered in the macroeconomics and finance literatures. These thick tails imply a lower tail index. However, as ν decreases, fewer moments exist, which could lead to infinite kurtosis and skewness, or even variance and expected value. This tradeoff between tail thickness and moment existence can prove burdensome for estimation of models like (3.31) through conventional methods. We propose the truncation based moment technique as a possible solution.

We consider three cases for our Monte Carlo experiments that are designed to capture distinct representations of the model in equation (3.31). In Case 1 we set $\rho = 0$ and estimate only the location parameter ω. In Case 2 we set $\omega = 0$ and estimate only the persistence parameter ρ. In Case 3 we estimate both the location and persistence parameters.

In each case we set the tail index of the innovation term large enough so that the expected value of the moment conditions used for estimation is finite, yet small enough so that they possess an infinite variance. As already discussed, the infinite variance problem precludes $\sqrt{T}$ Gaussian asymptotics for the estimators. Through Monte Carlo experiments, we endeavor to (i) illustrate the potential harm of using standard GMM techniques in this non-standard setting, and (ii) gauge the effectiveness of the Truncated GMM and Indirect Inference estimators at restoring Gaussian behavior.

Along with Gaussianity, we aim also to maintain, and if not improve, the finite sample accuracy of the estimators with these truncation techniques. Therefore, we explore a battery of sample statistics with which to asses the empirical goals of this paper. Near Gaussianity is measured through conventional metrics such as standard deviation (Std.Dev.), skewness (Skew), and kurtosis (Kurt). More formally, we test for Gaussianity with a Kolmogorov - Smirnov test. The null hypothesis of this test is a Gaussian distribution. The p-values (KSPval) are reported. We also report the maximum (Max) and minimum (Min) of the observed estimates. As for accuracy, bias is measured as $1/N \sum_{i=1}^{N} (\hat{\theta}_i - \theta^0)$, where N is the number of Monte Carlo paths, and θ may represent ω or ρ depending upon whether we are studying Case 1 or Case 2, respectively. We also provide the Mean Squared Error (MSE) and Mean Absolute Deviation (MAD), computed as $1/N \sum_{i=1}^{N} (\hat{\theta}_i - \theta^0)^2$, and $1/N \sum_{i=1}^{N} |\hat{\theta}_i - \theta^0|$, respectively. In Case 3, $\theta = [\omega \ \rho]'$, in which case all the sample statistics above are computed element-wise.

These statistics are computed for a standard GMM estimator $\hat{\theta}_{GMM}$, the Truncated GMM estimator $\hat{\theta}_{TGMM}$, and the Indirect Inference estimator $\hat{\theta}_{II}$. We also examine the Least Absolute Deviation estimator $\hat{\theta}_{LAD}$ as a benchmark for comparison.

A long literature supports the use of Least Absolute Deviations as a solution to the problem of parameter estimation in the face of heavy tails. Early work such as Gross and Steiger (1979),

and more recently Davis (1996), show that the LAD estimator is more accurate than the least squares estimator in the presence of infinite variance innovations.

It is often extolled that the robustness of the LAD estimator originates from the fact that it does not overweight extreme observations, as is done with the least squares estimator. Truncation techniques also seek to re-weight the influence of extreme observations. In this respect, the LAD and truncation approaches are similar in spirit. However, truncation identifies explicitly the extreme realizations and offers a much more flexible re-weighting scheme, than does LAD.

The LAD estimator is defined as follows. Consider the autoregressive process in equation (3.31), with $\theta = [\omega \ \rho]'$. Then

$$\hat{\theta}_{LAD} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T} |y_t - \omega - \rho y_{t-1}| \tag{3.32}$$

A popular method for computing this LAD estimator is as per Spyropoulous, Kiountouzis and Young (1973), and Abdelmalek (1974), who describe an approach similar to re-iterated weighted least squares. (See DasGupta and Mishra (2007) for an excellent review.)[7].

For notational convenience, define $\mathbf{Y}$ as a $(T \times 1)$ vector consisting of y_t for $t = 1, 2, ..., T$, and $\mathbf{X}$ as a $(T \times 2)$ vector ones appended by a vector containing y_{t-1} for $t = 1, 2, ..., T$. Moreover, define $\mathbf{u}$ as a $(T \times 1)$ vector consisting of u_t for $t = 1, 2, ..., T$, and θ as a (2×1) vector containing the location and persistence parameters. The vectorial version of equation (3.31) is then written as

$$\mathbf{Y} = \mathbf{X}\theta + \mathbf{u} \tag{3.33}$$

Begin by estimating equation (3.33) by least squares and obtain $\hat{\theta}_{LAD}$. Compute the estimated value $\hat{\mathbf{Y}} = \mathbf{X}\hat{\theta}$, and residual $\hat{\mathbf{u}} = \mathbf{Y} - \hat{\mathbf{Y}}$. Construct the matrix W where $\hat{W}_{st} = \frac{1}{|\hat{u}_{st}|}$ for $s = t$, else $\hat{W}_{st} = 0$, for $s, t = 1, 2, ..., T$. Then, update the estimate of θ via $\hat{\theta}_{LAD} = (\mathbf{X}'\hat{W}\mathbf{X})^{-1}\mathbf{X}'\hat{W}\mathbf{Y}$. Re-compute the estimated residual and repeat the process until the estimates of $\hat{\theta}_{LAD}$ converge to within some pre-defined tolerance level. We use this general method

[7]When applied in the field of Operations Research, this estimator is often computed using simplex methods such as the Barrodale-Roberts algorithm, the Bartels-Conn-Sinclair algorithm, Usow's L1-Fit algorithm, and the Bloomfield-Steiger algorithm.

for constructing the LAD estimator in the Cases below.

3.5.1 Case 1: Location

In Case 1 we estimate the location of an infinite variance process. We set $\rho = 0$, which reduces equation (3.31) to

$$y_t = \omega + u_t \tag{3.34}$$

The natural moment conditions are

$$\phi_t(\omega) = y_t - \omega \tag{3.35}$$

Their Winsorized counterparts are

$$\phi_t^{W,c} = \phi_t(\omega) Min \left[1, \frac{c}{|\phi_t(\omega)|}\right] \tag{3.36}$$

with sample mean $\bar{\phi}_T^{W,c}(\omega) = 1/T \sum_{t=1}^{T} \bar{\phi}_T^{W,c}(\omega)$.

The following Monte Carlo experiments center upon $N = 500$ paths of $y_t, t = 1, ..., T$ constructed via equation (3.34), each of length $T = 200$. The location parameter is set such that $\omega^0 = 2.0$. The innovation process is distributed $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. When the Indirect Inference approach is used, the binding function is approximated using five simulated paths, i.e. $H = 5$. Moreover, we assume the form of the innovation process is known. That is, the innovation process used to approximate the binding function is assumed to follow a Student's T with $\tilde{\nu}$ degrees of freedom, and we set $\tilde{\nu} = \nu^0$. We explore the possibility of mis-specifying the degrees of freedom parameter in the robustness checks at this end of this subsection. We do not explore, however, the use of an incorrect innovation distribution, i.e. something other than Student's T.

The tail index $\nu^0 = 1.5$ is chosen so that $E[u_t] = E[\phi_t(\omega)] < \infty$, while $Var[u_t] = Var[\phi_t(\omega)] = \infty$. This heavy tail behavior has a deleterious effect on the sample estimates of ω. The first column of Table 3.7 presents results for a standard GMM estimator without truncation, $\hat{\omega}_{GMM}$. This estimator is relatively accurate, but is skewed (-1.1) and exhibits a

large degree of excess kurtosis (33.1). The LAD estimator moderates the extreme behavior of the estimator, with skewness of -0.9 and kurtosis of 3.5, but at the expense of a loss of accuracy. For instance, the bias of $\hat{\omega}_{LAD}$ is -1.6. The truncation based estimators, on the other hand, are accurate, with a bias of 0.006 and 0.002 respectively. Moreover, $\hat{\omega}_{TGMM}$ and $\hat{\omega}_{II}$ exhibit near Gaussian behavior, as suggested by the average p-values from the Kolomogorov-Smirnov test of 0.54 for each estimator.

The last row of Table 3.7 provides a sense as to what extent the moment conditions are truncated. The entry $\%Trunc$ indicates the mean percentage of the realizations of the moment conditions that are truncated. For instance, a $\%Trunc = 50$ would imply that on average half of the realizations of $\phi_t(\omega)$ would be replaced by (c). In the Truncated GMM and Indirect Inference procedures, roughly 65% of the realizations are truncated. The actual truncation threshold corresponding to this percentage is 0.69. This value of the truncation threshold is found by minimizing an estimate of the variance matrix of the respective parameters.

Figure 3.7 illustrates the search for the variance minimizing truncation threshold. The $Var(\hat{\omega}_{II})$ reaches a minimum near 65% of the realized moment conditions being truncated, as already noted in Table 3.7. This moderately bowl-shaped pattern is the result of a tradeoff between the various components of the asymptotic variance of $\hat{\omega}_{II}$. Specifically, there is a tradeoff between the slope of the binding function, $\partial b/\partial\hat{\omega}_{II}$, and the variance of the truncated GMM estimator, $Var(\hat{\omega}_{TGMM})$. The bottom right panel of Figure 3.7 shows that $\partial b/\partial\hat{\omega}_{II} \approx 1$, which is to be expected in this simple case. That leaves the variance of the truncated GMM estimator as the driving force behind the $Var(\hat{\omega}_{II})$, as shown in the bottom left panel.

Since $\hat{\omega}_{TGMM}$ is a GMM estimator, we can see that it also consists of two components, $\partial\phi_t^{W,c}/\partial\hat{\omega}_{TGMM}$ and $Var[\phi_t^{W,c}(\hat{\omega}_{TGMM})]$. When few realizations are truncated, $\partial\phi_t^{W,c}/\partial\hat{\omega}_{TGMM}$ trivially is equal to -1, as depicted in the top left panel of Figure 3.7. As the percent of realizations truncated increases, i.e. as (c) decreases, fewer observations remain available for estimation. The moment conditions become unresponsive to changes in the underlying parameter.

The top right panel of Figure 3.7 depicts the heart of the infinite variance problem. When very little truncation is employed, the variance of the moment conditions is quite high. As

we increase the percent of realizations that are truncated, the variance falls to a level that is suitable for estimation and inference.

In summary, the top two panels of Figure 3.7 combine to generate the $Var(\hat{\omega}_{TGMM})$, which is depicted in the bottom left panel of that figure. In turn, the bottom two panels of Figure 3.7 combine to generate the pattern seen for the $Var(\hat{\omega}_{II})$, which is depicted in Figure 3.7.

We explore two ways in which to approximate the asymptotic variance of the estimators. We will elaborate on the variance of the Indirect Inference estimator here since its construction subsumes that of the Truncated GMM estimator. First, we approximate the various components of the theoretical variance matrix $\hat{\Sigma}_{II} = (1+1/H)\left[(\widehat{\partial b/\partial \omega})Var(\hat{\omega}_{TGMM})^{-1}(\widehat{\partial b/\partial \omega})\right]^{-1}$. The term $Var(\hat{\omega}_{TGMM})$ is approximated easily via standard GMM procedures, and the term $(\widehat{\partial b/\partial \omega})$ can be found through a straightforward application of finite differences. In a second approach, we use the Monte Carlo assessment of the variance. Specifically, $\hat{\Sigma}_{MC} = 1/N \sum_{i=1}^{N} (\hat{\omega}_{II,i} - \bar{\hat{\omega}}_{II})^2$.

The two methods yield quite similar results. For the experiment outlined in Table 3.7, the approximation of $Var(\hat{\omega}_{II})$ takes an average value of 0.0122, while the Monte Carlo assessment is 0.0103. Given this similarity of the estimates, we report results formed only with the Monte Carlo assessment for the duration of this paper.

The Indirect Inference approach relies upon an accurate guess of the form of the innovation process u_t. We assume that the econometrician knows the correct distribution of the innovation, but may have guessed incorrectly the tail index. This is a useful experiment to undertake not only because it justifies the Indirect Inference approach, but also because gauging the magnitude of tail behavior is an often lamented task in statistics and econometrics. Table 3.7 presents results for $\tilde{\nu} = [1.1, 1.3, 1.5, 1.7, 2.0]$, where $\nu^0 = 1.5$. In terms of both accuracy and dispersion, the $\hat{\omega}_{II}$ estimates are remarkably similar across the values of $\tilde{\nu}$. The standard deviation of the estimates is 0.10 and the KS p-value is 0.54 in all cases. Therefore, the Indirect Inference estimator seems rather robust to mis-specification within the similar.

Another practical concern is to assess the impact of using the truncation approaches when there is no infinite variance problem. We address this concern by simulating $u_t \sim N(0,1)$ and building $y_t, t = 1, ..., T$ according to equation (3.34). By construction the moment conditions

have finite variance. We compute the standard GMM and LAD estimators, and present the results in Table 3.7. As expected, $\hat{\omega}_{GMM}$ is quite accurate and Gaussian. The LAD estimates exhibit Gaussian behavior, but are not accurate. The Truncated GMM estimator is comparable to the standard GMM estimator in almost every respect. The finite sample biases are identical (0.005), and the skewness and kurtosis are nearly so. The Indirect Inference approach uses a simulator built from $\tilde{u}_t \sim StT(1.5)$. Again, the truncation technique is comparable to the standard GMM approach. We conclude that the TGMM and II estimators are robust even when they are applied to situations where they are not needed.

Notice that the threshold chosen from the TGMM and II estimators in Table 3.7 is only 1%. This suggests that in the face of moment conditions with finite variance, little truncation is required. Figure 3.7 expands on this idea. We compute $\hat{\omega}_{TGMM}$ for 500 paths, and chose the proper truncation threshold for each path by minimizing an estimate of the asymptotic variance. We repeat this experiment for tail indices $\nu^0 = [1.01, 1.02, ..., 2.0]$. Clearly, thicker tails require more truncation.

As noted in section 3.4.4, the Indirect Inference estimator is unnecessary from an asymptotic perspective. Table 3.7 compares the behavior of the Truncated GMM and Indirect Inference estimators as we increase the sample size. For a relatively small sample, $T = 200$, $\hat{\omega}_{TGMM}$ and $\hat{\omega}_{II}$ are similar. For example, the skewness, kurtosis, and standard deviation of the two estimators are identical. However, the bias, MSE, and MAD differ. As we increase the sample size to $T = 5,000$, a very rough proxy for an infinitely large sample, whatever differences there are virtually disappear. Every metric of accuracy and dispersion is identical to at least three decimal places. This finding supports the claim that Indirect Inference is unnecessary asymptotically when the data are symmetrically distributed data.

In a secondary analysis, we examine briefly the impact of asymmetric innovations upon the truncated estimators. We simulate a random variable $u_t^* \sim Pareto(k, a)$, where k is a scale parameter and a is a shape parameter. We then construct the de-meaned random variable $u_t = u_t^* - E[u_t^*]$. The zero-mean innovation term allows us to focus on the truncation as the only source of potential bias in the moment conditions, which is analogous to what we explored with the symmetric Student's T innovations. Notice that the higher order moments

of the Pareto random variable depend only on the shape parameter, and so are unaffected by this simple location shift.

The Pareto distribution shares a moment existence property similar to the Student't T. $E[u_t^{\gamma}] < \infty$ for any $0 < \gamma < a$. Therefore, in the following Monte Carlo experiments, we set the scale parameter $k^0 = 1$ and the shape parameter $a^0 = 1.5$, which allows $E[\phi_t(\rho)] < \infty$, but $Var[\phi_t(\rho)] = \infty$. This corresponds to the basic design of the symmetric innovations example.

Table 3.7 reveals that the standard GMM estimator once again is unable to produce accurate and approximately Gaussian estimators in the face of infinite variance. The LAD estimator is able to moderate somewhat the occurrence of extreme estimates. However, note that there are no claims that the LAD estimator should be applicable in such cases. The truncation based techniques are able to generate near Gaussian behavior, as suggested by the relatively high Kolomogorov-Smirnov p-values (0.15). However, the Truncated GMM estimator is inaccurate, as measured by a MSE of 1.8. The symmetric truncation scheme produces biased results in the face of asymmetrically distributed moment conditions. However, the Indirect Inference approach adjusts for the bias, reducing the MSE to just 0.007.

Unlike in the symmetric innovation case, this improvement provided by the Indirect Inference approach is not simply a finite sample phenomenon. Table 3.7 reveals that the relative performance of the TGMM and II estimators persists even when $T = 5000$.

Once again, the success of the Indirect Inference approach rests with its robustness to mis-specification of the simulator. In order to evaluate this robustness in the face of asymmetric innovations we conduct the following exercise. We assume that the innovation distribution is known, but the tail index used in the simulator is incorrect. Notice, this is similar in structure to the robustness check for the symmetric innovations example. However, varying the tail index in the Pareto distribution alters its skewness, unlike in the simple Student's T example. So, mis-specifying $\tilde{\nu}$ can be harmful to the estimator's accuracy. The structure of the experiment is identical to that presented in Table 3.7. We vary only the shape parameter used in the simulator so that $\tilde{a} = [1.1, 1.3, 1.7, 2.0]$.

Table 3.7 reveals that the dispersion of the estimates in unaffected by the tail index used in the simulator. The skewness and kurtosis remains steady across the various estimates, 0.34

and 3.18 respectively. However, the accuracy of the estimator is impacted greatly. The bias of $\hat{\omega}_{II}$ can be as high as 7.6 when $\tilde{a}$ used in the simulator is 1.1. In general, when the symmetry of the simulator does not resemble that of the DGP, the resulting estimates seem imprecise.

3.5.2 Case 2: Persistence

In Case 2 we estimate the persistence of a zero-mean autoregressive process. By setting $\omega = 0$, equation (3.31) reduces to

$$y_t = \rho y_{t-1} + u_t \tag{3.37}$$

The natural moment conditions are

$$\phi_t(\rho) = (y_t - \rho y_{t-1})\, y_{t-1} \tag{3.38}$$

Their Winsorized counterparts are

$$\phi_t^{W,c}(\rho) = \phi_t(\rho) Min\left[1, \frac{c}{|\phi_t(\rho)|}\right] \tag{3.39}$$

The following Monte Carlo experiments center upon $N = 500$ paths of $y_t, t = 1, ..., T$ constructed via equation (3.37), each of length $T = 200$. The persistence parameter is set such that $\rho^0 = 0.9$. The innovation process u_t follows $StT(\nu^0)$, where $\nu^0 = 2.1$. When the Indirect Inference approach is used, the binding function is approximated using twenty five simulated paths; i.e. $H = 25$. Moreover, we assume the simulator is correctly specified by a Student's T distribution and tail index $\tilde{\nu} = \nu^0$.

The moment conditions in equation (3.38) contain terms of the order u_t^2. By consequence, the variance of these moment conditions contain terms of the order u_t^4. Therefore, the $\nu^0 = 2.1$ implies that although $E[\phi_t(\rho)] < \infty$, the $Var[\phi_t(\rho)] = \infty$. This heavy tail behavior has a deleterious effect on the GMM estimator, although it is not quite as damaging as seen in Case 1.

Table 3.7 suggests that the GMM estimator $\hat{\rho}_{GMM}$ is relatively accurate, as indicated by an MSE of 0.001. However, the KSPval of 0.005 indicates that $\hat{\rho}_{GMM}$ does not exhibit

Gaussian behavior. The LAD estimator, on the other hand, seems to do a good job at restoring Gaussianity, as suggested by the KSPval of 0.08. The truncation based techniques also are able to generate accurate and nearly Gaussian estimators[8]. Their biases and standard deviations are less than that of the standard GMM estimator and the KSPval's are greater than 0.08.

As in Case 1, the proper truncation threshold is chosen by minimizing an estimate of the variance of ρ. Figure 3.7 illustrates that minimization across a range of potential truncation thresholds. The tradeoff between the slope of the binding function and the variance of $\hat{\omega}_{TGMM}$ suggests that, on average, only 18% of the realizations of the moment conditions should be truncated.

The details of this tradeoff are illustrated in Figure 3.7. Analogously to Figure 3.7, the bottom two panels capture the general tradeoff within the variance of $\hat{\omega}_{II}$, while the top tow panels focus on the variance of $\hat{\omega}_{TGMM}$.

The top left panel of Figure 3.7 suggests that as more realizations are truncated, the truncated moment condition becomes less informative about the underlying parameter. The top right panel illustrates the true motivation for this paper. When little truncated is employed, the variance of the moment conditions is relatively high. This reverses as the level of truncation increases.

The top two panels combine to yield the pattern of behavior for $Var(\hat{\rho}_{TGMM})$ seen in the bottom left panel of Figure 3.7. A high variance estimate can be subdued with truncation. The bottom right panel might best be read from right to left along the x-axis. As we truncate an increasing proportion of the realizations of the moment conditions, the binding function becomes more informative. This process continues until we reach a point where further truncation is irrelevant, which is signalled by the flat spot beginning near 30% truncation.

Table 3.7 examines the sample performance of the II estimator at truncation thresholds

[8]Notice that the bias of $\hat{\rho}_{TGMM}$ is smaller than that of $\hat{\rho}_{II}$; -0.002 versus -0.006. However, as we increase the number of replications in the simulator from $H = 25$ to $H = 100$, the bias of $\hat{\rho}_{II}$ drops to 0.001. This supports the finding of Gourieroux, Renault and Touzi (1999) that Indirect Inference can offer a finite sample advantage over standard GMM techniques if enough replications are used to provide an accurate approximation of the binding function.

other than 18%. The standard deviation of the estimators produced from the various truncation thresholds ranges only from 0.0226 to 0.0294. However, the dispersion of the estimators can vary significantly. For truncation thresholds near 18%, the KSPval is relatively high, usually near 0.09. However, for very low truncation thresholds, the KSPval can be as low as 0.013. In all, it appears that any moderate level of truncation will produce estimators with better accuracy and dispersion properties than the standard GMM estimator without truncation.

As suggested in Case 1, the proper amount of truncation depends on the tail index of the innovation process. As illustrated in Figure 3.7, thicker tails require more truncation.

The Indirect Inference estimator appears relatively robust to mis-specification in the simulator. We replicate the experiment outlined in Table 3.7, while varying the tail index used in the simulator for the II estimator. We consider $\tilde{\nu} = [2.05, 2.15, 4]$, when $\nu^0 = 2.1$. Table 3.7 reveals that $\hat{\rho}_{II}$ is relatively accurate and nearly Gaussian, regardless of the tail index used in the simulator. For example, the standard deviation is always 0.023, and the kurtosis ranges only from 3.47 to 3.62.

The TGMM and II estimators also are robust to mistakenly using these truncation techniques in situations when they are not needed. We simulate $u_t \sim N(0,1)$ and compute $\hat{\rho}_{GMM}$ and $\hat{\rho}_{LAD}$. As expected, the standard GMM estimator is accurate, with a bias of only -0.009, and nearly Gaussian, with a KSPval of 0.15. The LAD estimator does not perform quite as well, with a KSPval of 0.024. The TGMM estimator is comparable to the GMM estimator in terms of accuracy and dispersion. The bias of $\hat{\rho}_{TGMM}$ is only -0.008, and its KSPval is 0.12. The Indirect Inference estimator also produces results that are comparable to the GMM estimator, regardless of whether we use $\tilde{u}_t \sim StT(2.1)$ or $\tilde{u}_t \sim StT(4.0)$.

The near Gaussianity of the Indirect Inference estimator appears to diminish as the level of persistence in the process increases. In Table 3.7 we examine $\rho^0 = [0.2, 0.4, 0.8, 0.9, 0.95, 0.99]$. The KSPval ranges from 0.66 for $\rho^0 = 0.2$ to 0.00 for $\rho^0 = 0.99$. This suggests that the truncation techniques, as outlined in this paper, may not be appropriate for highly persistent processes. We leave this question as future research.

In the face of symmetric innovations, the TGMM and II should be asymptotically equivalent. Table 3.7 corroborates this claim. For a small sample of $T = 200$, $\hat{\rho}_{TGMM}$ and $\hat{\rho}_{II}$ are

somewhat similar. The bias, for instance, is -.006 and -0.002, respectively. However, as the sample size grows to $T = 5000$, the two estimators are virtually indistinguishable.

The truncation based estimators also are successful at estimating the persistence parameter in the face of asymmetric innovations. With a scale $k^0 = 1$ and a shape parameter $a^0 = 3.1$, we simulate u_t according to a Pareto distribution, and construct $N = 500$ paths of $y_t, t = 1, ..., T$ according to equation (3.37), each of length $T = 200$. With $a^0 = 3.1$, the first three moments of the innovation term exist. Notice however that $\phi_t(\rho)$ contains terms of the order u_t^2. Consequently $E[\phi_t(\rho)] < \infty$, but $Var[\phi_t(\rho)] = \infty$.

Table 3.7 shows that the $\hat{\rho}_{GMM}$ and $\hat{\rho}_{LAD}$ estimators are relatively accurate, but exhibit only moderate degrees of Gaussian behavior. The KSPval's are 0.016 and 0.068, respectively. The TGMM estimator, on the other hand, exhibits near Gaussianity, with KSPval of 0.281. The II estimator is a bit less accurate, bias of 0.019, and exhibits less evidence of Gaussianity, KSPval of 0.108. Still, each of the truncation techniques seems to be an overall improvement over the untruncated GMM estimator.

The differences between the TGMM and II estimators evaporates as the sample size grows. Table 3.7 reveals that, for example, the difference in the bias of the estimators shrinks from 0.008 to 0.001 when the sample size is large. Similarly, the difference in the KSPval's shrinks from 0.17 to 0.0. This result suggests that estimation of the persistence parameter is not impacted by the symmetry of the innovation term.

The efficacy of the Indirect Inference estimator relies upon it's robustness to mis-specifications of the simulator. We conduct two exercises to assess this robustness. First, we repeat the experiment described in Table 3.7, while altering only the shape parameter used in the simulator. Specifically, we simulate $u_t \sim Pareto(1, a^0)$, where $\tilde{a} = [2.1, 2.8, 3.4, 4.0]$, and $a^0 = 3.1$. Table 3.7 reveals that the distribution of the II estimator seems little effected by the mis-specification at $\tilde{a}$. The standard deviations of the estimates ranges from 0.017 to 0.019, and the kurtosis ranges from 3.16 to 3.24. However, accuracy appears to deteriorate when the skewness of the distribution implied by the simulator differs from the DGP. For example, the bias can be as low as 0.008 when $\tilde{a} = 2.1$ and as high as 0.024 when $\tilde{a} = 4.0$.

A second robustness exercise assumes that the true Pareto innovation distribution is mistaken for a Gamma distribution. This seems like a reasonable error given the distributions' similar behavior. The Gamma is a two parameter distribution, $G(k, \alpha)$, where k is a scale parameter and α is a shape parameter. Since the higher order moments of the Gamma depend only on the shape parameter, as they do with the Pareto, we again set $k^0 = 1$. All that remains is knowledge of α, which can be inferred by observing the empirical skewness of y_t.

Given the stationary AR(1) structure in equation (3.37), it is straightforward to show that the skewness of y_t is $\gamma^y = \gamma^u \frac{1-\rho^3}{(1-\rho^2)^{3/2}}$. This implies that $\gamma^u = \gamma^y \frac{(1-\rho^2)^{3/2}}{1-\rho^3}$, the estimate of which consists of two pieces. First, we can approximate γ^y with the empirical skewness of the observed series such that $\hat{\gamma}^y = \frac{1/T \sum_{t=1}^{T} (y_t - \bar{y})^3}{[1/T \sum_{t=1}^{T} (y_t - \bar{y})^2]^{3/2}}$. Second, is an estimate of ρ, which of course, is the parameter of interest, and thus unknown. Fortunately, we have available the Truncated GMM estimator as a first step approximation to ρ. Therefore $\hat{\gamma}^u = \hat{\gamma}^y[(1-\hat{\rho}^2_{TGMM})^{3/2}/(1-\hat{\rho}^3_{TGMM})]$

We also know that the skewness of a Gamma distributed random variable is $2/\sqrt{\alpha}$. By setting this theoretical level of skewness to the target skewness $\hat{\gamma}^u$, we can solve for α. Following this approach implies that the innovations in the simulator of the Indirect Inference approach should come from a Gamma distribution with $\tilde{\alpha} = (2/\hat{\gamma}^u)^2$.

Table 3.7 presents estimates of the II estimator with the skewness matching approach. The $\hat{\rho}_{II}$ is less accurate than the estimator found with a correctly specified simulator or the TGMM estimator. The bias, for instance, is -0.043 , as compared to -0.011 for $\hat{\rho}_{TGMM}$. However, $\hat{\rho}_{II}$ exhibits a strong degree of Gaussian behavior, as indicated by the KSPval of 0.5.

These findings suggest that the Indirect Inference approach is relatively robust to a misspecified simulator as long as we are able to mimic reasonably well the skewness of the true innovation process.

3.5.3 Case 3: Location and Persistence

In Case 3 we estimate the location and persistence of an AR(1) process. The model of interest is displayed in equation (3.31).

The natural moment conditions are

$$\phi_t(\theta) = \begin{bmatrix} \phi_{1,t}(\theta) \\ \phi_{2,t}(\theta) \end{bmatrix} = \begin{bmatrix} y_t - \omega - \rho y_{t-1} \\ (y_t - \omega - \rho y_{t-1}) y_{t-1} \end{bmatrix} \tag{3.40}$$

where $\theta = [\omega \ \rho]'$ and $\phi_{t,i}(\theta)$ is row (i) of the time (t) vector of moment conditions.

The following Monte Carlo experiment is similar in structure to Cases 1 and 2. We generate $N = 500$ paths of $y_t, t = 1, ..., T$ according to equation (3.31), each of length $T = 200$. The location and persistence parameters are set at $\omega^0 = 2.0$ and $\rho^0 = 0.9$. The innovation term is distributed $u_t \sim StT(\nu^0)$, where $\nu^0 = 2.1$. When the Indirect Inference approach is used, the binding function is approximated with $H = 5$ simulations. Moreover, we assume no misspecification of the innovation term used in the simulator, so that $\tilde{\nu} = \nu^0 = 2.1$.

The tail index of the innovation term $\nu^0 = 2.1$ impacts the two moment conditions differently. Both have finite mean, but only the first of the two has finite variance. $E[\phi_{1,t}(\rho)] < \infty$, $E[\phi_{2,t}(\rho)] < \infty$, and $Var[\phi_{1,t}(\rho)] < \infty$, but $Var[\phi_{2,t}(\rho)] = \infty$. This suggests that we may want to customize the truncation scheme. In fact, the presence of multiple moment conditions presents the opportunity for multiple truncation thresholds.

The purpose of this subsection is to illustrate how multiple moment conditions and multiple thresholds can be incorporated into the Truncated GMM and Indirect Inference strategies.

First, we may use a single threshold for truncation, where (c) is a scalar and the truncation is based upon the norm of the time (t) realization of the moment conditions. Specifically,

$$\phi_t^{W,c}(\theta) = \phi_t(\theta) Min \left[1, \frac{c}{||\phi_t(\theta)||}\right] \tag{3.41}$$

Alternatively, we could use multiple truncation thresholds as described in section 3.4.3 and detailed here as follows:

$$\phi_t^{W,c}(\theta) = \begin{bmatrix} \phi_{1,t}^{W,c}(\theta) \\ \\ \phi_{2,t}^{W,c}(\theta) \end{bmatrix} = \begin{bmatrix} \phi_{1,t}(\theta) Min \left[1, \frac{c(1)}{|\phi_{1,t}(\theta)|}\right] \\ \\ \phi_{2,t}(\theta) Min \left[1, \frac{c(2)}{|\phi_{2,t}(\theta)|}\right] \end{bmatrix} \tag{3.42}$$

In the particular setting of Case 3, only $\phi_{2,t}(\theta)$ requires truncation. Therefore, we set $c(1) = \infty$, which implies no truncation and choose $c(2)$ by minimizing the variance of θ, exactly as done in the previous sub-sections.

Table 3.7 consists of four panels. The top two panels capture results where a single truncation threshold is used, as described in equation (3.41). The top left panel presents results for ω, and the top right panel captures results for ρ. The bottom two panels capture results where only the second of the moment conditions is truncated, as described in equation (3.42).

The GMM estimates of ω are heavily biased (0.403), but do exhibit a degree of Gaussian behavior, with a KSPval of 0.047. The LAD estimator offers some improvement in terms of accuracy, with a bias of 0.145, but little in the way of near Gaussianity.

In general, the truncation based estimators offer an advantage over GMM without truncation and LAD regardless of the truncation scheme. However, some important differences are apparent.

The biases of $\hat{\omega}_{TGMM}$ and $\hat{\omega}_{II}$ are smaller than those of $\hat{\omega}_{GMM}$ and $\hat{\omega}_{LAD}$ in both the single and separate threshold schemes. Moreover, the standard deviations of the truncation based estimators in the single threshold scheme are about 0.475, while they are 0.756 and 0.499 for the GMM and LAD estimators, respectively. However, this advantage is reversed in the multiple threshold scheme, where the standard deviations are about 1.2 for the truncation based methods.

In terms of near Gaussianity, $\hat{\omega}_{TGMM}$ and $\hat{\omega}_{II}$ have a clear advantage over $\hat{\omega}_{GMM}$ and $\hat{\omega}_{LAD}$ in both truncation schemes. The KSPval = 0.047 and 0.055 for the GMM and LAD estimators, but are always at least 0.21 for TGMM and II in the single threshold scheme, and are as high as 0.39 in the multiple threshold scheme.

The findings for ρ are similar. The GMM estimator is reasonably accurate, and exhibits some Gaussian behavior. The LAD estimator offers no improvement. The truncation based estimators offer an improvement in accuracy when a single threshold is used, but are less accurate when multiple thresholds are used. However, the evidence of Gaussianity is strong. The KSPval for $\hat{\rho}_{II}$ is above 0.6 within both truncation schemes.

A key distinguishing feature of the truncation schemes is the significant difference in the

percent of realizations that are truncated. In the single threshold scheme, 65% of the realizations of the moment conditions need to be truncated in order to minimize the variance of θ. Yet only 1% of the realizations need to be truncated in the multiple threshold scheme.

3.6 Conclusion

In this paper, we offer a way in which to conduct standard GMM estimation and inference when the variance of the moment conditions is infinite. The Truncated GMM ($\hat{\theta}_{TGMM}$) and Indirect Inference ($\hat{\theta}_{II}$) estimators we present here truncate the moment conditions, which ensures they have a finite variance. Standard asymptotic theory then applies. Furthermore, we advocate choosing the optimal truncation threshold by minimizing a sample estimate of the variance for the parameters of interest.

Monte Carlo work suggests that the truncation-based estimation offers an improvement over standard GMM estimation without truncation, and in many cases, over the Least Absolute Deviation (LAD) estimator as well. We find that the Indirect Inference estimator offers a finite sample advantage over the Truncated GMM estimator, consistent with the findings of Gourieroux, Renault and Touzi (1999). Asymptotically, that advantage appears to depend upon the symmetry of the observed process. When the data are symmetric, Truncated GMM and Indirect Inference are virtually identical in our simulation work. However, when the data are asymmetric, the Indirect Inference estimator may still offer an advantage asymptotically. This is found to be true in the case of estimating a location parameter, but not in the case of estimating an autocorrelation parameter. We find also that the Indirect Inference estimator is relatively robust to mis-specification of the simulator used to approximate the binding function.

Possible extensions of this technique include exploring alternative truncation schemes, such as trimming or smoothing. In addition, alternative objective functions, such as those produced by Continuous Updated GMM and Empirical Likelihood, may be worthy of exploration.

Although the applications for this approach abound, we aim next to explore the infinite variance problem within an ARMA-GARCH model. The Self-Weighted QMLE of Ling (2007) and the Modified QMLE of Lange, Rahbek and Jensen (2007) are natural contenders in this

setting.

3.7 Tables & Figures

Table 3.1: Case 1: Symmetric Innovations - Small Sample

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II
Bias	-0.002	-1.615	0.006	0.004
Std.Dev.	0.791	0.146	0.100	0.100
Skew	-1.066	-0.939	-0.090	-0.090
Kurt	33.121	3.472	2.738	2.738
Max	8.906	0.670	2.304	2.302
Min	-4.512	-0.010	1.716	1.715
KSPval	0.000	0.001	0.543	0.543
MSE	0.624	2.630	0.010	0.010
MAD	0.400	1.615	0.082	0.081
%Trunc	0	0	65	65

Table 3.2: Case 1: Symmetric Innovations - Large Sample

(Case 1) *$\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$T = 200$		$T = 5000$	
	TGMM	II	TGMM	II
Bias	0.006	0.004	0.000	0.000
Std	0.100	0.100	0.019	0.019
Skew	-0.090	-0.090	-0.118	-0.118
Kurt	2.738	2.738	3.150	3.150
Max	2.304	2.302	2.054	2.054
Min	1.716	1.715	1.943	1.943
KSPval	0.543	0.543	0.872	0.872
MSE	0.010	0.010	0.000	0.000
MAD	0.082	0.081	0.015	0.015
%Trunc	65	65	60	60

Table 3.3: Case 1: Mis-specifying Simulator Under Symmetry

(Case 1) *$T = 200$, $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. Results for the Indirect Inference approach are reported. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$\tilde{\nu}$				
	1.1	1.3	1.5	1.7	2.0
Bias	0.010	0.002	0.004	0.015	0.010
Std	0.100	0.100	0.100	0.100	0.100
Skew	-0.090	-0.090	-0.090	-0.090	-0.090
Kurt	2.738	2.738	2.738	2.738	2.738
Max	2.308	2.300	2.302	2.313	2.308
Min	1.721	1.713	1.715	1.726	1.720
KSPval	0.543	0.543	0.543	0.543	0.543
MSE	0.010	0.010	0.010	0.010	0.010
MAD	0.082	0.081	0.081	0.083	0.082
%Trunc	65	65	65	65	65

Table 3.4: Case 1: Truncation Works Even If Not Needed

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim N(0,1)$. $\tilde{u}_t \sim StT(\tilde{\nu})$, $\tilde{\nu} = 1.5$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II
Bias	0.005	-1.198	0.005	-0.022
Std	0.070	0.036	0.071	0.071
Skew	-0.202	-0.223	-0.217	-0.217
Kurt	2.979	2.979	2.996	2.996
Max	2.184	0.896	2.185	2.157
Min	1.747	0.686	1.747	1.719
KSPval	0.951	0.629	0.927	0.927
MSE	0.005	1.437	0.005	0.005
MAD	0.057	1.198	0.057	0.059
%Trunc	0	0	1	1

Table 3.5: Case 1: Asymmetric Innovations - Small Sample

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim Pareto(k, a)$, $k^0 = 1$ and $a^0 = 1.5$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{a} = a^0$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II
Bias	0.046	-1.895	-1.352	-0.034
Std	1.531	0.068	0.075	0.075
Skew	11.130	0.541	0.344	0.344
Kurt	173.752	2.984	3.177	3.177
Max	28.050	0.344	0.927	2.244
Min	0.956	0.000	0.448	1.765
KSPval	0.000	0.050	0.145	0.145
MSE	2.342	3.594	1.833	0.007
MAD	0.603	1.895	1.352	0.068
%Trunc	0	0	55	55

Table 3.6: Case 1: Asymmetric Innovations - Large Sample

(Case 1) $\omega^0 = 2.0$*;* $H = 5$*;* $N = 500$*. Innovations:* $u_t \sim Pareto(k, a)$*, where* $k^0 = 1$ *and* $a^0 = 1.5$*. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e.* $\tilde{a} = a^0$*. TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$T = 200$		$T = 5000$	
	TGMM	II	TGMM	II
Bias	-1.350	-0.030	-1.360	-0.020
Std	0.080	0.080	0.020	0.020
Skew	0.340	0.340	-0.150	-0.150
Kurt	3.180	3.180	2.830	2.830
Max	0.930	2.240	0.690	2.030
Min	0.450	1.770	0.590	1.940
KSPval	0.150	0.150	0.900	0.900
MSE	1.830	0.010	1.850	0.000
MAD	1.350	0.070	1.360	0.020
%Trunc	55	55	55	55

Table 3.7: Case 1: Mis-specifying Simulator Under Asymmetry

(Case 1) $\omega^0 = 2.0$*;* $H = 5$*;* $N = 500$*; Innovations:* $u_t \sim Pareto(k^0, a^0)$*,* $k^0 = 1$*, and* $a^0 = 3.1$*. The innovations used in the simulator of the Indirect Inference approach are distributed* $\tilde{u}_t \sim Pareto(k^0, \tilde{a})$*. TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

				$\tilde{a}$				
	GMM	LAD	TGMM	1.1	1.3	1.5	1.7	2.0
Bias	0.046	-1.895	-1.352	7.595	1.146	-0.034	-0.510	-0.830
Std	1.531	0.068	0.075	0.075	0.075	0.075	0.075	0.075
Skew	11.130	0.541	0.344	0.344	0.344	0.344	0.344	0.344
Kurt	173.752	2.984	3.177	3.177	3.177	3.177	3.177	3.177
Max	28.050	0.344	0.927	9.873	3.424	2.244	1.768	1.448
Min	0.956	0.000	0.448	9.395	2.946	1.765	1.290	0.970
KSPval	0.000	0.050	0.145	0.145	0.145	0.145	0.145	0.145
MSE	2.342	3.594	1.833	57.693	1.319	0.007	0.266	0.695
MAD	0.603	1.895	1.352	7.595	1.146	0.068	0.510	0.830
%Trunc	0	0	55	55	55	55	55	55

Table 3.8: Case 2: Symmetric Innovations - Small Sample

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II
Bias	-0.008	-0.003	-0.002	-0.006
Std	0.031	0.022	0.023	0.023
Skew	-0.784	-0.543	-0.529	-0.501
Kurt	3.717	3.863	3.586	3.554
Max	0.954	0.959	0.957	0.954
Min	0.755	0.804	0.807	0.804
KSPval	0.005	0.080	0.087	0.088
MSE	0.001	0.000	0.001	0.001
MAD	0.024	0.017	0.018	0.018
%Trunc	0	0	17	18

Table 3.9: Case 2: Symmetric Innovations - Large Sample

(Case 2) *$\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, where $\nu^0 = 2.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$T = 200$		$T = 5000$	
	TGMM	II	TGMM	II
Bias	-0.006	-0.002	0.000	0.000
Std	0.023	0.023	0.000	0.000
Skew	-0.501	-0.539	-0.040	-0.040
Kurt	3.554	3.592	2.760	2.770
Max	0.954	0.957	0.910	0.910
Min	0.804	0.807	0.890	0.890
KSPval	0.088	0.099	0.990	0.990
MSE	0.001	0.001	0.000	0.000
MAD	0.018	0.018	0.000	0.000
%Trunc	18.000	16.000	9.000	9.000

Table 3.10: Case 2: Mis-specifying Simulator Under Symmetry

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. Results for the Simulation Based Truncated GMM approach are reported. The innovations in the simulator are distributed $u_t \sim StT(\tilde{\nu})$. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$\tilde{\nu}$			
	2.05	2.1	2.15	4
Bias	-0.002	-0.006	-0.004	-0.004
Std	0.023	0.023	0.023	0.023
Skew	-0.546	-0.501	-0.425	-0.447
Kurt	3.620	3.554	3.472	3.515
Max	0.958	0.954	0.959	0.959
Min	0.806	0.804	0.806	0.805
KSPval	0.066	0.088	0.210	0.165
MSE	0.001	0.001	0.001	0.001
MAD	0.018	0.018	0.018	0.018
%Trunc	17	18	19	17

Table 3.11: Case 2: Effectiveness Across Truncation Thresholds

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. Results for the Indirect Inference approach are reported. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	%Trunc.								
	1	5	10	16	18	20	30	50	90
Bias	-0.0047	-0.0052	-0.0055	-0.0058	-0.0058	-0.0059	-0.0058	-0.0059	-0.0054
Std	0.0277	0.0242	0.0231	0.0226	0.0226	0.0227	0.0232	0.0246	0.0294
Skew	-0.7471	-0.588	-0.5577	-0.5118	-0.5012	-0.4912	-0.4943	-0.4664	-0.7153
Kurt	4.0391	3.359	3.5083	3.5449	3.5537	3.5525	3.4489	3.2316	5.1077
Max	0.9484	0.9459	0.952	0.9542	0.9543	0.9539	0.9527	0.9558	1.0031
Min	0.764	0.8137	0.8069	0.8044	0.8044	0.804	0.8111	0.8091	0.731
KSPval	0.013	0.0139	0.2167	0.1113	0.0878	0.0936	0.2236	0.2176	0.0834
MSE	0.0008	0.0006	0.0006	0.0005	0.0005	0.0005	0.0006	0.0006	0.0009
MAD	0.0212	0.019	0.0183	0.0179	0.0179	0.018	0.0184	0.0196	0.0225

Table 3.12: Case 2: Effectiveness Across Persistence Parameters

(Case 2) *$T = 200$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, where $\nu^0 = 2.1$. Results for the Indirect Inference approach are reported. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	ρ^0					
	0.2	0.4	0.8	0.9	0.95	0.99
Bias	-0.001	-0.006	-0.005	0.001	-0.005	-0.003
Std	0.053	0.050	0.031	0.023	0.017	0.012
Skew	-0.213	-0.345	-0.380	-0.527	-0.737	-1.536
Kurt	3.463	3.480	3.278	3.576	3.923	6.146
Max	0.370	0.529	0.873	0.960	0.987	1.010
Min	0.009	0.217	0.677	0.810	0.876	0.930
KSPval	0.658	0.333	0.189	0.068	0.001	0.000
MSE	0.003	0.002	0.001	0.001	0.000	0.000
MAD	0.041	0.039	0.025	0.018	0.013	0.008
%Trunc	5	5	14	18	20	30

Table 3.13: Case 2: Truncation Works Even If Not Needed

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim N(0, 1)$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. The innovation term in the simulator of the Indirect Inference approach is distributed $u_t \sim StT(\tilde{\nu})$. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II($\tilde{\nu} = 2.1$)	II($\tilde{\nu} = 4.0$)
Bias	-0.009	-0.01	-0.008	-0.009	-0.007
Std	0.033	0.041	0.033	0.034	0.034
Skew	-0.502	-0.398	-0.488	-0.443	-0.492
Kurt	3.181	2.901	3.146	3.070	3.169
Max	0.969	0.995	0.973	0.976	0.977
Min	0.773	0.76	0.777	0.774	0.772
KSPval	0.153	0.024	0.123	0.143	0.099
MSE	0.001	0.002	0.001	0.001	0.001
MAD	0.027	0.034	0.027	0.028	0.028
%Trunc	0	0	1	2	3

Table 3.14: Case 2: Asymmetric Innovations - Small Sample

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 250$. Innovations: $u_t \sim Pareto(k^0, a^0)$, $k^0 = 1$ and $a^0 = 3.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{a} = a^0$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation estimator; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	GMM	LAD	TGMM	II
Bias	-0.016	-0.009	-0.011	0.019
Std	0.031	0.033	0.021	0.017
Skew	-0.474	-0.316	-0.232	-0.293
Kurt	3.163	3.352	3.006	3.193
Max	0.949	0.992	0.948	0.968
Min	0.774	0.798	0.823	0.862
KSPval	0.016	0.068	0.281	0.108
MSE	0.001	0.001	0.001	0.001
MAD	0.027	0.027	0.019	0.022
%Trunc	0	0	7	7

Table 3.15: Case 2: Asymmetric Innovations - Large Sample

(Case 2) *$\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim Pareto(k, a)$, $k^0 = 1$ and $a^0 = 3.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{a} = a^0$. TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

	$T = 200$		$T = 5000$	
	TGMM	II	TGMM	II
Bias	-0.011	0.019	-0.005	-0.004
Std	0.021	0.017	0.004	0.004
Skew	-0.232	-0.293	-0.024	-0.029
Kurt	3.006	3.193	2.721	2.705
Max	0.948	0.968	0.906	0.907
Min	0.823	0.862	0.882	0.884
KSPval	0.281	0.108	0.500	0.500
MSE	0.001	0.001	0.000	0.000
MAD	0.019	0.022	0.005	0.005
%Trunc	7	7	6	6

Table 3.16: Case 2: Mis-specifying Tail Index Under Asymmetry

(Case 2) *$\rho^0 = 0.9$; $H = 25$; $N = 500$; Innovations: $u_t \sim Pareto(k^0, a^0)$, $k^0 = 1$, and $a^0 = 3.1$. The innovations in the simulator of the Indirect Inference approach are distributed as $\tilde{u}_t \sim Pareto(k^0, \tilde{a})$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated.*

				$\tilde{a}$				
	GMM	LAD	TGMM	2.1	2.8	3.1	3.4	4.0
Bias	-0.016	-0.009	-0.011	0.008	0.017	0.019	0.021	0.024
Std	0.031	0.033	0.021	0.019	0.017	0.017	0.017	0.017
Skew	-0.474	-0.316	-0.232	-0.298	-0.301	-0.293	-0.311	-0.320
Kurt	3.163	3.352	3.006	3.171	3.164	3.193	3.208	3.235
Max	0.949	0.992	0.948	0.960	0.966	0.968	0.969	0.971
Min	0.774	0.798	0.823	0.847	0.859	0.862	0.864	0.868
KSpval	0.016	0.068	0.281	0.106	0.103	0.108	0.084	0.072
MSE	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001
MAD	0.027	0.027	0.019	0.016	0.020	0.022	0.023	0.025
%Trunc	0	0	7	7	7	7	7	7

Table 3.17: Case 2: Mis-specifying Distribution Under Asymmetry

(Case 2) *$\rho^0 = 0.9$; $H = 25$; $N = 500$; Innovations: $u_t \sim Pareto(k^0, a^0)$, $k^0 = 1$, and $a^0 = 3.1$. The innovations in the simulator of the Indirect Inference approach are distributed as $\tilde{u}_t \sim Gamma(k^0, \tilde{\alpha})$, where $\tilde{\alpha}$ is inferred from the empirical skewness of y_t. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated*

	GMM	LAD	TGMM	II
Bias	-0.016	-0.009	-0.011	-0.043
Std	0.031	0.033	0.021	0.021
Skew	-0.474	-0.316	-0.232	-0.114
Kurt	3.163	3.352	3.006	3.191
Max	0.949	0.992	0.948	0.922
Min	0.774	0.798	0.823	0.799
KSpval	0.016	0.068	0.281	0.500
MSE	0.001	0.001	0.001	0.002
MAD	0.027	0.027	0.019	0.043
%Trunc	0	0	7	3

Table 3.18: Case 3: Symmetric Innovations

(Case 3) *$T = 200$; $\rho^0 = 0.9$; $\omega^0 = 2.0$; $H = 25$; $N = 250$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. The innovations used in the simulator of the Indirect Inference approach are correctly specified, both in terms of the distribution and the degrees of freedom; i.e. $\tilde{\nu} = \nu^0$. GMM = GMM estimator without truncation; LAD = Least Absolute Deviation; TGMM = Truncated GMM; II = Indirect Inference. KSPval = p-value from Kolmogorov - Smirnov test for Gaussianity. %Trunc. = percent of realizations of moment conditions that are truncated. "Single Threshold" implies that we use the norm of the moment conditions in the Winsorizing mechanism. "Multiple Threshold" implies that we truncate only the second of the two moment conditions.*

	$\hat{\omega}$				$\hat{\rho}$			
Single Threshold	GMM	LAD	TGMM	II	GMM	LAD	TGMM	II
Bias	0.403	0.145	0.174	0.100	-0.020	-0.007	-0.009	-0.005
Std	0.756	0.499	0.470	0.475	0.035	0.024	0.023	0.023
Skew	0.838	0.582	0.609	0.616	-0.707	-0.496	-0.520	-0.522
Kurt	3.826	3.374	3.737	3.755	3.461	3.277	3.564	3.582
Max	5.149	3.727	3.832	3.789	0.954	0.963	0.945	0.950
Min	0.879	0.798	1.113	1.027	0.750	0.822	0.814	0.816
KSPval	0.047	0.055	0.253	0.211	0.236	0.014	0.575	0.626
MSE	0.731	0.269	0.250	0.235	0.002	0.001	0.001	0.001
MAD	0.634	0.392	0.377	0.368	0.030	0.019	0.019	0.018
%Trunc	0	0	65	65	0	0	65	65
Multiple Threshold	GMM	LAD	TGMM	II	GMM	LAD	TGMM	II
Bias	0.403	0.145	0.266	0.152	-0.020	-0.007	-0.012	-0.006
Std	0.756	0.499	1.124	1.160	0.035	0.024	0.050	0.053
Skew	0.838	0.582	0.528	0.441	-0.707	-0.496	-0.433	-0.305
Kurt	3.826	3.374	2.670	2.586	3.461	3.277	2.585	2.511
Max	5.149	3.727	5.316	5.211	0.954	0.963	1.008	1.013
Min	0.879	0.798	-0.141	-0.075	0.750	0.822	0.760	0.764
KSPval	0.047	0.055	0.231	0.390	0.236	0.014	0.300	0.614
MSE	0.731	0.269	1.329	1.363	0.002	0.001	0.003	0.003
MAD	0.634	0.392	0.911	0.937	0.030	0.019	0.041	0.043
%Trunc	0	0	1	1	0	0	1	1

Figure 3.1: Case 1: Choosing The Truncation Threshold

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. x-axis: %Trunc. = percent of realizations of moment conditions that are truncated. y-axis: $Var[\hat{\omega}_{II}]$*

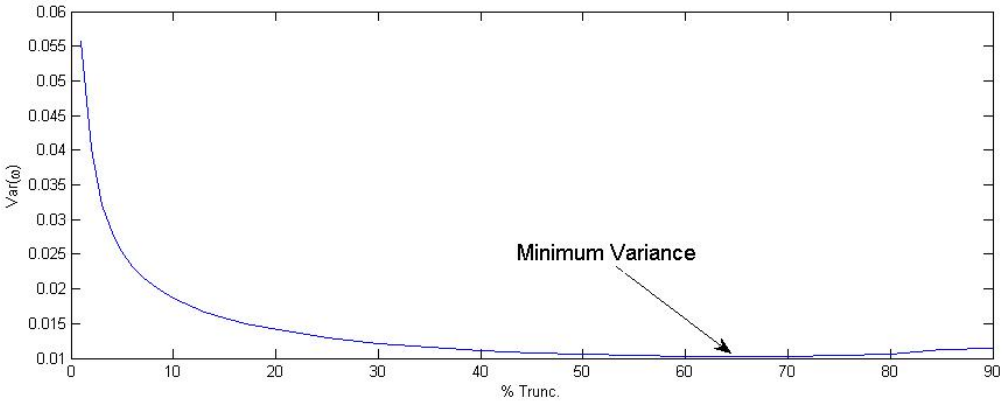


Figure 3.2: Case 1: Dissecting $Var(\hat{\theta}_{II})$

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 1.5$. x-axis: %Trunc. = percent of realizations of moment conditions that are truncated. y-axis labels are as follows; Top Left Panel: $\partial\phi^{W,c}(\hat{\omega}_{TGMM})/\partial\omega$. Top Right Panel: $Var[\phi_t^{W,c}(\hat{\omega}_{TGMM})]$. Bottom Left Panel: $Var[\hat{\omega}_{TGMM}]$. Bottom Right Panel: $\partial b(\hat{\omega}_{II})/\partial\omega$.*

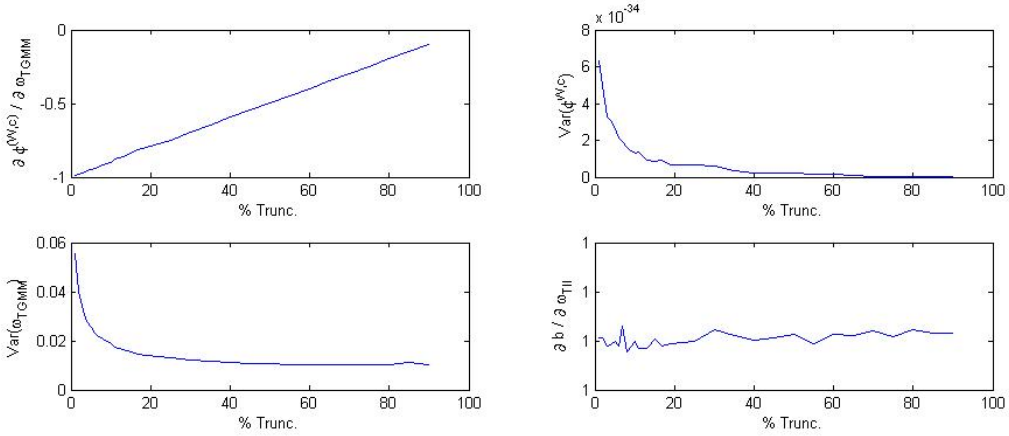


Figure 3.3: Case 1: Truncation Threshold And Tail Index

(Case 1) *$T = 200$; $\omega^0 = 2.0$; $H = 5$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$. Estimates computed via the Truncated GMM approach. x-axis: %Trunc. = percent of realizations of moment conditions that are truncated. y-axis: ν^0.*

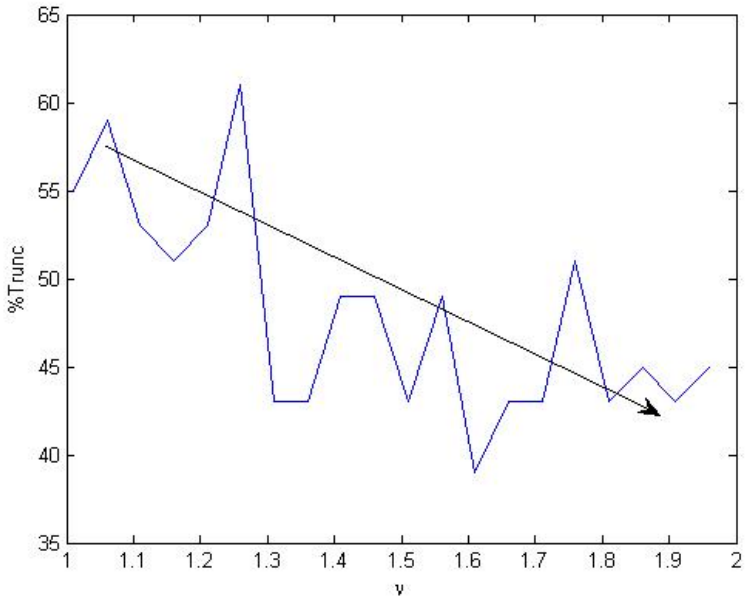


Figure 3.4: Case 2: Choosing The Truncation Threshold

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. x-axis: $\%Trunc.$ = percent of realizations of moment conditions that are truncated. y-axis: $Var[\hat{\rho}_{II}]$.*

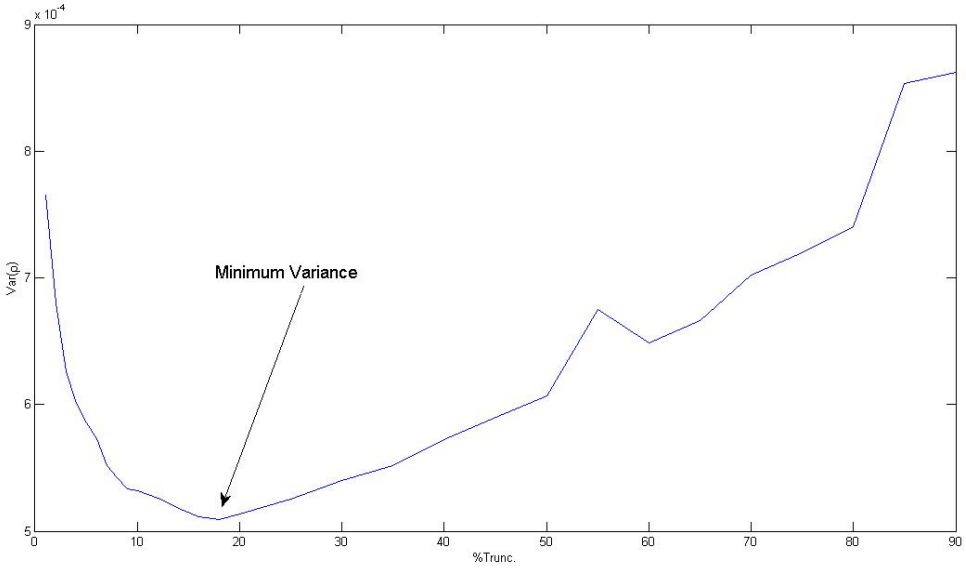


Figure 3.5: Case 2: Dissecting $Var(\hat{\theta}_{II})$

(Case 2) *$T = 200$; $\rho^0 = 9.0$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$, $\nu^0 = 2.1$. x-axis: $\%Trunc.$ = percent of realizations of moment conditions that are truncated. y-axis labels are as follows; Top Left Panel: $\partial\phi^{W,c}(\hat{\rho}_{TGMM})/\partial\rho$. Top Right Panel: $Var[\phi_t^{W,c}(\hat{\rho}_{TGMM})]$. Bottom Left Panel: $Var[\hat{\rho}_{TGMM}]$. Bottom Right Panel: $\partial b(\hat{\rho}_{II})/\partial\rho$.*

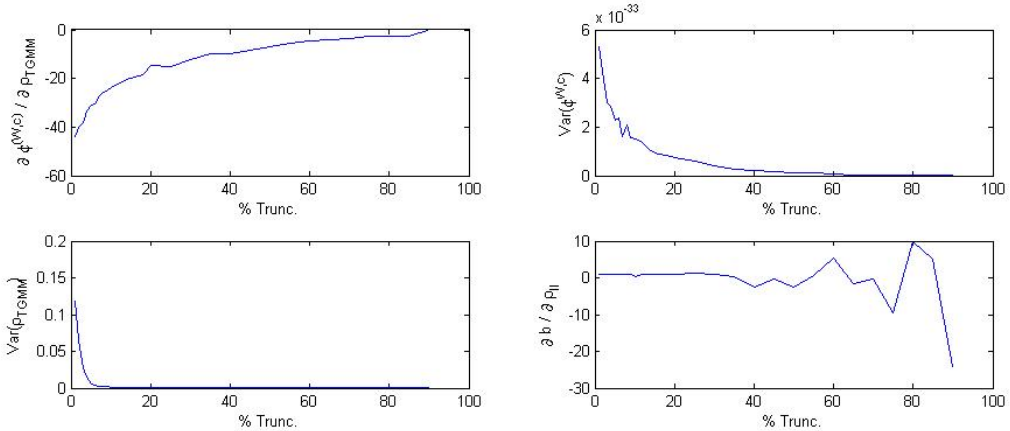
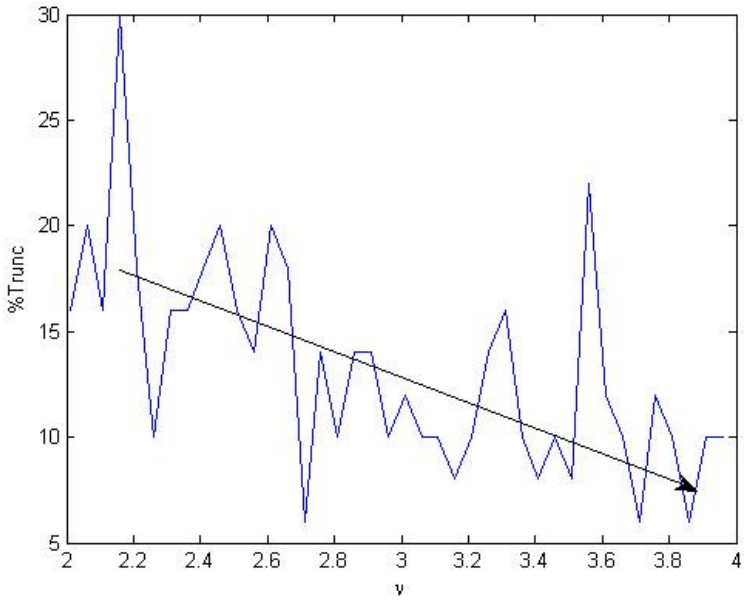


Figure 3.6: Case 2: Truncation Threshold And Tail Index

(Case 2) *$T = 200$; $\rho^0 = 0.9$; $H = 25$; $N = 500$. Innovations: $u_t \sim StT(\nu^0)$. Estimates computed via the Truncated GMM approach. x-axis: %Trunc. = percent of realizations of moment conditions that are truncated. y-axis: ν^0.*



Chapter 4

A Moment Based Test of GARCH Against Stochastic Volatility

4.1 Introduction

Modeling the unobserved volatility of asset returns is a well explored line of research in finance and econometrics. The benchmarks in this endeavor are Generalized Autoregressive Conditional Heteroscedasticity (GARCH) as per Engle (1982) and Bollerslev (1986), and Stochastic Volatility (SV) as per Taylor (1982) and others. Although there are important differences in these models, ranging from theoretical to practical implementation, little has been done in the way of formalizing a specification test in order to distinguish them.

A notable reference is Kobayashi and Shi (2005), which designs a specification test of EGARCH against its Stochastic Volatility counterpart. In addition, Franses, Leij and Paap (2008) present a test of a GARCH model with Student's T innovations against its Stochastic Volatility counterpart.

Each of these previous efforts are likelihood based and thus parametric in nature. In this paper, we offer a semi-parametric approach by proposing a specification test of GARCH against Stochastic Volatility within a GMM framework.

This chapter is organized as follows. In section 4.2, we develop an original way to nest a GARCH(1,1) model within a Stochastic Volatility model. This nesting will allow us to test for the GARCH(1,1) specification through a GMM-based inference sketched in section 4.3. A

short scale Monte Carlo study of GMM performance in this setting is presented in section 4.4.

4.2 Nesting GARCH within SV

A common wisdom is that SV models introduce an additional exogenous random component in volatility dynamics, in contrast to GARCH models, wherein volatility is a deterministic function of past returns. Even though this was implicit in the seminal paper by Rosenberg (1972), it is often forgotten that simply adding noise to a GARCH(1,1) equation is a convenient way to think about stochastic volatility. Of course, a sensible framework for testing GARCH against a more general SV model must introduce this noise with some discipline.

Consider a SV model:

$$y_{t+1} = \sigma_t \varepsilon_{t+1} \tag{4.1}$$

where ε_{t+1} is a strong white noise with zero mean and σ_t is the volatility process. Linear SV dynamics, as introduced by Andersen (1994) under the name SR-SARV(1), and developed in a semi-parametric way for the purpose of GMM inference by Meddahi and Renault (2004), can be written as:

$$\sigma_t^2 = c + \gamma \sigma_{t-1}^2 + \xi_t \tag{4.2}$$

where $0 < \gamma < 1$ and ξ_t is a weak white noise. Meddahi and Renault (2004) note that the GARCH(1,1) model is a particular case of the SR-SARV(1), corresponding to a specific form of the SV-innovation process ξ_t.

The GARCH(1,1) model (written with SV notations) is given by

$$\sigma_t^2 = \omega + \alpha y_t^2 + \beta \sigma_{t-1}^2 \tag{4.3}$$

with $0 \leq \alpha, 0 \leq \beta, \alpha + \beta < 1$. Translating this GARCH(1,1) into a SR-SARV(1) requires only that we relax the restriction on the form of the innovation process. By defining $\gamma = \alpha + \beta$ and $\xi_t = \alpha[y_t^2 - \sigma_{t-1}^2]$, the GARCH(1,1) can be re-written as

$$\sigma_t^2 = \omega + \gamma \sigma_{t-1}^2 + \xi_t$$

As suggested by Lamoureux and Lastrapes (1990), we can build any SR-SARV(1) model simply by imagining that the intercept ω in the GARCH equation 4.3 actually is a time varying non-negative white noise process ω_t. Our discipline will amount to impose that this additional source of randomness does not modify the order of magnitude of the volatility process, as defined by 4.2. The conditional variance must remain proportional to $[1-\gamma^2]^{-1}$. However, note that volatility is generally highly persistent, and so $[1-\gamma^2] = (1+\gamma)(1-\gamma) \approx 2(1-\gamma)$. Therefore, we get an equivalent and simpler parametrization by assuming throughout that

$$Var(\omega_t) = \frac{\eta^2}{1-\gamma}$$

If we let $c = E[\omega_t]$, then our GARCH(1,1) model with time varying intercept can be written as a quite general SR-SARV(1) model.

$$\begin{aligned} \sigma_t^2 &= \omega_t + \alpha y_t^2 + \beta \sigma_{t-1}^2 && (4.4) \\ &= c + \gamma \sigma_{t-1}^2 + \xi_t && (4.5) \end{aligned}$$

where $\gamma = \alpha + \beta$ and $\xi_t = \alpha[y_t^2 - \sigma_{t-1}^2] + (\omega_t - c)$.

In addition, by adapting an argument due to Franses, Leij and Paap (2008), it is easy to check that equation (4.5) defines a non-negative process for any specification of the non-negative white noise ω_t. We can also show that the process $k_t = [\sigma_t^2 - \omega_t]$ is non-negative, by noting that from equation (4.5) it can be decomposed recursively as $k_t = \alpha y_t^2 + \beta k_{t-1} + \beta \omega_{t-1}$.

Within this framework, the hypothesis of a GARCH(1,1) process is tantamount to testing

$$H_0 : \eta^2 = 0$$

Franses, Leij and Paap (2008) propose a likelihood ratio test of this null hypothesis by assuming that the standardized innovation ε_t is normal or Student's-T, while the non-negative white noise is governed by a log-normal distribution. By contrast, in the next section we extend the GMM inference put forth in Meddahi and Renault (2004) to define a semi-parametric testing procedure.

4.3 GMM Inference

Consider the Stochastic Volatility process

$$y_{t+1} = \sigma_t \varepsilon_{t+1} \tag{4.6}$$

$$\sigma_t^2 = \omega + \alpha y_t^2 + \beta \sigma_{t-1}^2 + u_t \tag{4.7}$$

with $E[u_t] = 0$, $Var[u_t] = \frac{\eta^2}{1-\gamma}$, $E[\varepsilon_{t+1}] = 0$, $E\left[\varepsilon_{t+1}^2\right] = 1$, $E\left[\varepsilon_{t+1}^4\right] = \kappa$, and define $\gamma = \alpha + \beta$. Furthermore, assume $\{u_t\}$ is i.i.d. and independent of ε_t.

Our goal is to test $H_0 : \eta^2 = 0$. If this is the case, then equation (4.7) reduces to $\sigma_t^2 = \omega + \alpha y_t^2 + \beta \sigma_{t-1}^2$, which, along with equation (4.6), constitutes a standard GARCH(1,1) model.

The key implications of this model are

$$E\left[\sigma_t^2 | I_{t-1}\right] = \omega + \gamma \sigma_{t-1}^2 \tag{4.8}$$

$$Var\left[\sigma_t^2 | I_{t-1}\right] = \frac{\eta^2}{1-\gamma} + c\sigma_{t-1}^4 \tag{4.9}$$

where $c = \alpha^2(\kappa - 1)$.

From equations (4.8) and (4.9) we deduce

$$E\left[\sigma_{t-1}^4 | I_{t-1}\right] = \omega^2 + \frac{\eta^2}{1-\gamma} + 2\omega\gamma\sigma_{t-1}^2 + (c+\gamma)\sigma_{t-1}^4 \tag{4.10}$$

Since $y_{t+1} = \sigma_t \varepsilon_{t+1}$, we know that

$$E\left[y_t^4 | I_{t-2}\right] = \kappa \sigma_{t-1}^4$$

$$E\left[y_{t+1}^4 | I_{t-1}\right] = \kappa E\left[\sigma_t^4 | I_{t-1}\right]$$

By subtracting these two equations and using equation (4.10) we get

$$E\left[y_{t+1}^4 - (\gamma^2 + c) y_t^4 | I_{t-1}\right] = \kappa \left[\omega^2 + \frac{\eta^2}{1-\gamma}\right] + 2\omega\gamma\kappa\sigma_{t-1}^2 \tag{4.11}$$

From (4.11) and (4.8), we deduce

$$E\left[y_{t+1}^4 - (\gamma^2 + c)y_t^4 | I_{t-2}\right] = \kappa\left[\omega^2 + \frac{\eta^2}{1-\gamma} + 2\omega^2\gamma\right] + 2\omega\gamma^2\kappa\sigma_{t-2}^2$$

But by (4.11)

$$E\left[y_{t+1}^4 - (\gamma^2 + c)y_t^4 | I_{t-2}\right] = \kappa\left[\omega^2 + \frac{\eta^2}{1-\gamma}\right] + 2\omega\gamma\kappa\sigma_{t-2}^2 \tag{4.12}$$

Therefore, by subtracting these last two equations, with the later premultiplied by γ, we get

$$E\left[y_{t+1}^4 - (\gamma + \gamma^2 + c)y_t^4 + \gamma(\gamma^2 + c)y_{t-1}^4 - \xi | I_{t-2}\right] = 0 \tag{4.13}$$

where $\xi = \kappa(\omega^2 + \frac{\eta^2}{1-\gamma})(1-\gamma) + 2\omega^2\gamma\kappa = \kappa\eta^2 + \kappa\omega^2(1+\gamma)$.

This implies that

$$\frac{\xi}{\kappa} = \eta^2 + \omega^2(1+\gamma) \tag{4.14}$$

From (4.8) and the assumed structure of y_t, we know that

$$\begin{aligned} E\left[y_{t+1}^2 | I_{t-1}\right] &= E\left[\sigma_{t-1}^2 | I_{t-1}\right] \\ &= \omega + \gamma\sigma_{t-1}^2 \\ &= \omega + \gamma E\left[y_t^2 | I_{t-1}\right] \end{aligned}$$

This provides

$$E\left[y_{t+1}^2 - \omega - \gamma y_{t-1}^2 | I_{t-1}\right] = 0 \tag{4.15}$$

Equations (4.15) and (4.13) provide the moment conditions for GMM estimation, which we will use in a sequential manner.

In the first GMM problem, we estimate ω and γ via

$$E\left[y_{t+1}^2 - \gamma y_t^2 - \omega | I_{t-1}\right] = 0 \tag{4.16}$$

which provides $\tilde{\omega}$ and $\tilde{\gamma}$.

In the second GMM problem we estimate $\theta = [\omega, \gamma, c, \xi]'$ via

$$\phi_t(\theta) = \begin{bmatrix} \phi 1_t(\theta) \\ \phi 2_t(\theta) \end{bmatrix} = \begin{bmatrix} E\left[y_{t+1}^2 - \gamma y_t^2 - \omega | I_{t-1}\right] \\ E\left[y_{t+1}^4 - \gamma y_t^4 - c(y_t^4 - \tilde{\gamma} y_{t-1}^4) - \tilde{\gamma}^2 y_t^4 + \tilde{\gamma}^2 y_t^4 + \tilde{\gamma}^3 y_{t-1}^4 - \xi \big| I_{t-2}\right] \end{bmatrix} = 0 \tag{4.17}$$

where $\xi = \kappa\eta^2 - \tilde{\omega}^2(1+\tilde{\gamma})$, which provides $\hat{\omega}$, $\hat{\gamma}$, $\hat{c}$, and $\hat{\xi}$.

The reason for the sequential procedure is to remain linear in the parameters in each GMM problem. As detailed by Gourieroux, Monfort and Renault (1996), the weighting matrix in a setting such as this might be quite cumbersome. We can simplify matters greatly by utilizing the two-step procedure outlined above as long as the estimating equations are linear in the parameters.

From (4.14) we know that $\eta^2 = \xi/\kappa - \omega^2(1+\gamma)$. With $c = \alpha^2(\kappa - 1)$, we find that

$$\eta^2 = \frac{\xi}{1 + \frac{c}{\alpha^2}} - \omega^2(1+\gamma)$$

With parameter estimates in hand, we can test the hypothesis $\eta^2 = 0$, based on the asymptotic distribution of the parameter vector $\left[\hat{\omega}, \hat{\gamma}, \hat{c}, \hat{\xi}, \hat{\alpha}\right]'$, where α may be estimated by QML.

An alternative to this direct test on η^2 would be a simple Hausman test on α. Assuming that we have a Gaussian structure, we know that $\kappa = 3$, which implies that $\hat{\alpha} = \frac{\hat{c}}{2}^{1/2}$.

Notice that this estimator of α is consistent even if $\eta \neq 0$. By contrast, we also can estimate α by MLE. Under the null of a GARCH(1,1) model, this estimator is consistent and efficient. The null hypothesis for the Hausman test is $H_0 : p\lim(\hat{\alpha} - \hat{\alpha}^0) = 0$. Under the null,

$$\sqrt{T}(\hat{\alpha} - \hat{\alpha}^0) \rightarrow N(0, Var(\sqrt{T}\hat{\alpha}) - Var(\sqrt{T}\hat{\alpha}^0))$$

Note that much work remains in order to formalize properly these testing procedures. Several issues need to be addressed, including the boundary on these maintained hypotheses and the potentially non-standard distributions of the estimators.

4.4 Monte Carlo

The inference procedures described above rely upon the standard $\sqrt{T}$ Gaussian asymptotics that typically accompany GMM estimators. These asymptotics are dependent upon the existence of the variance of the moment conditions. Unfortunately, the variance of the moment conditions outlined in equation (4.17) may be infinite, and thus require special attention.

The existence of the variance of the moment conditions is important for several reasons, first and foremost of which pertains to the construction of the optimal weighting matrix. As put forth by Hansen (1982), an optimal choice of the weighting matrix in the GMM objective function is the inverse of the variance matrix of the moment conditions. If that variance matrix doesn't exist, this concept of the optimal weighting matrix makes little sense.

Notice that the moment conditions in equation (4.17) contain terms of the order y_t^4, implying that their variance contains terms of the order y_t^8. Finite eighth moments of a GARCH-type process is unlikely in practice. Figure 4.5 illustrates the frontier of existence for the 4^{th} and 8^{th} moments of y_t, as derived by Bollerslev (1986) with Gaussian innovation terms ε_{t+1}. All points to the left of the curve correspond to combinations of α and β where the indicated moment is finite.

We conduct a simple Monte Carlo exercise to illustrate the potential harmful effect the infinite variance of the moment conditions can have upon the parameter estimates.

We begin by generating $N = 250$ paths of $y_t, t = 1, ..., T$ for $T = 2,500$ according to equations (4.6) and (4.7). In addition, $\varepsilon_t \sim N(0,1)$, $\omega^0 = .001$, $\alpha^0 = 0.10$, and $\beta^0 = 0.85$, which implies $\gamma^0 = 0.95$. Moreover, $u_t = 0 \ \forall t$. Note that one can reasonably expect to encounter these values of α and β in practice when studying daily equity return volatility.

In the first GMM problem we use equations (4.16) to estimate ω and γ. Notice that these moment conditions contain terms of the order y_t^2, and so their variance contains terms of the order y_t^4. According to the existence conditions depicted in Figure 4.5, the variance of these moment conditions is finite. Table 4.5 reports the GMM estimates of ω and γ, as well as MLE estimates for comparison. We compute the Monte Carlo assessment of various sample statistics, such as mean, skewness, MSE, and others.

The GMM estimates of ω and γ are reasonably accurate and are close to the MLE values. For instance, the MSE of $\hat{\omega}$ estimated via GMM is 0.000 and its standard deviation is only 0.002. The GMM estimates of γ are not quite as good, with a MSE value of 0.016 and standard deviation of 0.122. However, even these estimates of γ remain reasonable.

Table 4.5 displays estimates of c and η^2 for the moment conditions in (4.17). The GMM estimates of η^2, which is zero by definition, are quite accurate. The estimates of c, on the other hand, are inaccurate (MSE = 2.009) and highly variable (Std.Dev. = 0.455).

The infinite variance problem clearly has a detrimental impact upon the estimates of c. We propose using the truncated moment technique detailed in Aguilar, Hill and Renault (2008) (Chapter 4 of this dissertation) as a remedy.

We begin by truncating each of the $i = 1, 2$ moment conditions in equation (4.17) using a Winsorizing technique

$$\phi_{i,t}^{W,b(i)}(\theta) = \phi_{i,t}(\theta) Min\left[1, \frac{b(i)}{|\phi_{i,t}(\theta)|}\right] \tag{4.18}$$

where $b(i)$ is a vector of positive components, known as truncation thresholds.

As Aguilar, Hill and Renault (2008) indicate, if the moment conditions are asymmetrically distributed, the truncation scheme may introduce an asymptotic bias. As such, the authors advocate using an Indirect Inference approach to adjust for the bias during estimation.

Of course, many issues remain, not least of which is the appropriateness of the simulator used during the Indirect Inference method. However, we are hopeful that as per Aguilar, Hill and Renault (2008) and Dridi, Guay and Renault (2007), the simulator will be robust to mis-specification.

4.5 Tables & Figures

Table 4.1: Monte Carlo - Moments With Finite Variance

y_t follows a GARCH(1,1) simulated according to equations (4.6) and (4.7); $\omega^0 = 0.001$, $\alpha^0 = 0.1$, $\beta^0 = 0.85$, $\gamma^0 = 0.95$, $\epsilon_t \sim N(0,1)$, $u_t = 0\ \forall t$. Sample Size= 2,500; # of Simulated Paths = 250. Results reported for MLE and GMM estimates using moment conditions in equation (4.16).

	ω		γ	
	MLE	GMM	MLE	GMM
Mean	0.001	0.002	0.946	0.912
Std	0.000	0.002	0.015	0.122
Skew	0.611	-0.042	-0.718	0.032
Kurt	3.693	3.535	3.697	3.412
Max	0.002	0.008	0.976	1.350
Min	0.000	-0.007	0.891	0.587
MSE	0.000	0.000	0.000	0.016
MAD	0.000	0.002	0.012	0.099

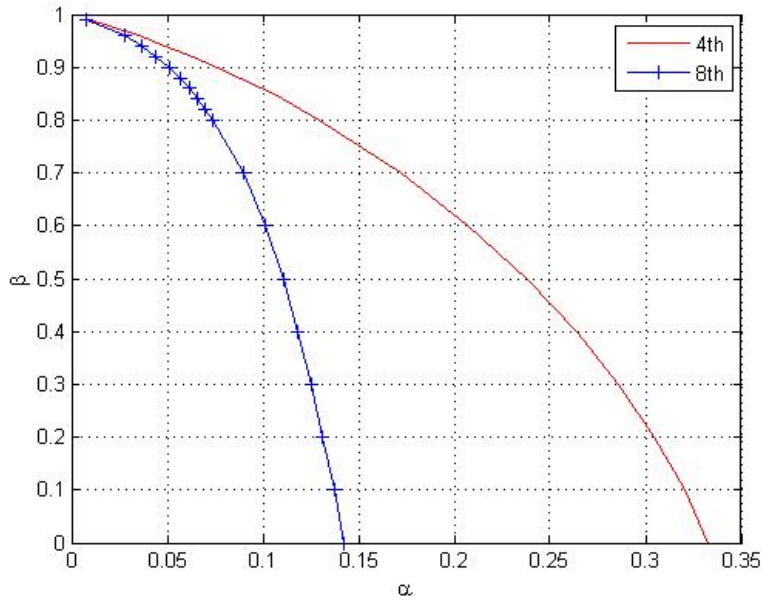
Table 4.2: Monte Carlo - Moments With Infinite Variance

y_t follows a GARCH(1,1) simulated according to equations (4.6) and (4.7); $\omega^0 = 0.001$, $\alpha^0 = 0.1$, $\beta^0 = 0.85$, $\gamma^0 = 0.95$, $\epsilon_t \sim N(0,1)$, $u_t = 0\ \forall t$, $(\eta^2)^0 = 0.0$, $c^0 = 2(\alpha^0)^2 = 0.02$. Sample Size= 2,500; # of Simulated Paths = 250. Results reported for GMM estimates using the moment conditions in equation (4.17).

	η^2	c
Mean	0.000	-1.323
Std	0.000	0.455
Skew	0.641	0.217
Kurt	3.766	2.509
Max	0.000	-0.175
Min	0.000	-2.279
MSE	0.000	2.009
MAD	0.000	1.343

Figure 4.1: Conditions for Moment Existence

Consider the GARCH(1,1) y_t process in equation (4.6) and (4.7) where $\epsilon_t \sim N(0,1)$ and $u_t = 0\ \forall t$. The figure illustrates the frontier for the existence of moments four and eight of y_t as derived by Bollerslev (1986) across various combinations of α and β. All points to the left of the curve represent finite moments for the indicated moment.



Appendix A

The Portfolio Allocation Problem

I follow Han (2006) in designing the portfolio allocation problem faced by the investor. Let $\mu_{p,t+1}$ and $\sigma^2_{p,t+1}$ denote the conditional mean and variance of the portfolio returns $y_{p,t+1}$. The goal is to minimize volatility subject to a target rate of return (μ^*_p), The investor allows for short sales and the existence of a risk free rate (r_f). All other variables are as defined in the body of this paper. The objective function and solution for the portfolio weights ($\mathbf{w}$) are below:

Minimum Volatility:

$$\min_{\mathbf{w}_t}\{\sigma^2_{p,t+1} = \mathbf{w}_t'\Sigma_{t|t+1}\mathbf{w}_t\}$$

subject to:

$$\mathbf{w}_t'\mu_{t+1|t} + (1 - \mathbf{w}_t'1)r_f = \mu^*_p$$

Define $\kappa_t = (\mu_{t+1|t} - 1r_f)'\Sigma_{t+1|t}(\mu_{t+1|t} - 1r_f)$. The portfolio weights are:

$$\mathbf{w}_t = \Sigma^{-1}_{t+1|t}(\mu_{t+1|t} - 1r_f)\frac{\mu^*_p - r_f}{\kappa_t}$$

Appendix B

Hausman Test

Along with our Monte Carlo experiments detailed in section 3.5, we may examine the robustness of our estimator by comparing it to another that we believe to be consistent. Fortunately, we have in mind a consistent benchmark for comparison. The estimator $\hat{\theta}_{HR}$ designed by Hill and Renault (2008), henceforth HR, is a consistent truncation based estimator of θ. The HR estimator is similar in spirit to our own in that the authors truncate the moment conditions within a GMM framework in order to restore Gaussian asymptotics. However, the truncation threshold used by HR increases with the sample size, whereas ours remains fixed. Therefore, as T gets large, their threshold will tend to grow toward infinity, which would eventually fail to eliminate the infinite variance problem. The key to the HR approach is that they are able to find the proper speed of convergence, and so are able to establish a consistent and asymptotically normal estimator. Specifically, $\sqrt{g_T}(\hat{\theta}_{HR} - \omega^0) \xrightarrow{d} N(0, \Sigma_{HR})$, for some $g_T = o(T)$.

Although $\hat{\theta}_{HR}$ has the advantage of being consistent without the worry of a binding function, it may be less efficient than $\hat{\theta}_{II}$ because of its potentially slower rate of convergence.

We can inform our choice among these two estimators via a Hausman test. The null and alternative hypotheses are:

$$\begin{aligned} H_0 : \operatorname*{plim}_{T\to\infty}(\hat{\theta}_{HR} - \hat{\theta}_{II}) &= 0 \\ H_a : \operatorname*{plim}_{T\to\infty}(\hat{\theta}_{HR} - \hat{\theta}_{II}) &\neq 0 \end{aligned}$$

Under the null,

$$\sqrt{g_T}(\hat{\theta}_{HR} - \hat{\theta}_{II}) = \sqrt{g_T}(\hat{\theta}_{HR} - \theta^0) + \sqrt{g_T}(\theta^0 - \hat{\theta}_{II}) \tag{B.1}$$

The second term on the right hand side of the (B.1) converges to zero. Therefore,

$$\sqrt{g_T}(\hat{\theta}_{HR} - \hat{\theta}_{II}) \xrightarrow{d} N(0, \Sigma_{HR})$$

Under the alternative, the second term on the right hand side of (B.1) diverges. Consequently, $\sqrt{g}_T(\hat{\theta}_{HR} - \hat{\theta}_{II})$ also does not converge.

Therefore, we can form an appropriate Hausman test statistic as

$$\zeta = g_T(\hat{\theta}_{HR} - \hat{\theta}_{II})'(\Sigma_{HR})^{-1}(\hat{\theta}_{HR} - \hat{\theta}_{II}) \tag{B.2}$$

If $\zeta < \chi^2_1$, we fail to reject the null, suggesting that the truncated estimator is indeed consistent. Faced with two consistent estimators, we choose the one with the faster rate of convergence; i.e. $\hat{\theta}_{II}$.

Bibliography

Abdelmalek, N. 1974. "On the Discrete Linear L^1 Approximation and L^1 Soluation of Over-Determined Linear Equations." *Journal of Approximation Theory* 11:35–53.

Aguilar, M., J. Hill and E. Renault. 2008. "Indirect Inference for Moment Equations with Infinite Variance." Working Paper - UNC-CH.

Aguilar, O. and M. West. 2000. "Bayesian dynamic factors models and variance matrix discounting for portfolio allocation." *Journal of Business and Economic Statistics* 18:338–357.

Alexander, C. 2001. "A Primer on the Orthogonal GARCH Model." Working Paper - ISMA Centre.

Andersen, T.G. 1994. "Stochastic autoregressive volatility: a framework for volatility modeling." *Mathematical Finance* 4:75–102.

Andrews, D. 1994. *Empirical Process Methods in Econometrics*. in: Engle, R., and McFadden, D. (Eds.), "Handbook of Econometrics 4", North-Holland.

Asai, M., M. McAleer and J. Yu. 2006. "Multivariate Stochastic Volatility: A Review." *Econometric Reviews* 25:145–157.

Bickel, P. 1965. "On Some Robust Estimates of Location." *The Annals of Mathematical Statistics* 36(3):847–858.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–327.

Carrasco, M. 2007. "A regularization approach to the many instruments problem." Working Paper - University of Montreal.

Connor, G., R.A. Korajczyk and O. Linton. 2006. "The common and specific components of dynamic volatility." *Journal of Econometrics* 132:231–255.

Cox, D. 1966. "The null distribution of the first serial correlation coefficient." *Biometrika* 53:623–626.

DasGupta, M. and SK. Mishra. 2007. "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review." Working Paper - NEHU.

Davis, R. 1996. "Gauss-Newton and M-estimation for ARMA processes with infinite variance." *Stochastic Processes and their Applications* 63:75–95.

Davis, R., K. Knight and J. Liu. 1992. "M-estimation for autoregressions with infinite variance." *Stochastic Processes and Their Applications* 40:145–180.

Diebold, F.X. and M. Nerlove. 1989. "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model." *Journal of Applied Econometrics* 4:1–22.

Dovonon, P. 2006. "Conditionally Heteroskedastic Factor Models with Skewness and Leverage Effects." Working Paper - University of Montreal.

Doz, C. and E. Renault. 2006. "Factor Stochastic Volatility in Mean Models: A GMM Approach." *Econometric Reviews* 25:275–309.

Dridi, R., A. Guay and E. Renault. 2007. "Indirect Inference and calibration of dynamic stochastic general equilibrium models." *Journal of Econometrics* 136:397–430.

Eichenbaum, M., L.P. Hansen and K.J. Singleton. 1998. "A Time Series of Representative Agent Models of Consumption and Leisure Choice under Uncertainty." *Quarterly Journal of Economics* 103:51–78.

Engle, R. 2001. "Dynamic Conditional Correlation - A simple class of multivariate GARCH models." *Journal of Business and Economic Statistics* .

Engle, R. and K. Sheppard. 2001. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." Working Paper WP-8554 - NBER.

Engle, R.F. 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation." *Econometrica* 64:813–836.

Engle, R.F. and S. Kozicki. 1993. "Testing for Common Features." *Journal of Business and Economic Statistics* 11:369–395.

Fiorentini, G. and E. Sentana. 2001. "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models." *Journal of Econometrics* 102:143–164.

Fiorentini, G., E. Sentana and N. Shephard. 2004. "Likelihood-Based Estimation of Latent Generalized ARCH Structures." *Econometrica* 72.

Fleming, J., C. Kirby and B. Ostdiek. 2001. "The Economic Value of Volatility Timing." *Journal of Finance* 56:329–352.

Francq, C. and J. Zakoian. 2004. "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes." *Bernouli* 10(4):605–637.

Franses, P., M. Leij and R. Paap. 2008. "A Simple Test for GARCH against a Stochastic Volatility Model." *Journal of Financial Econometrics* 6(3):291–306.

French, K., G. Schwert and R. Stambaugh. 1987. "Expected Stock Returns and Volatility." *Journal of Financial Economics* 19(1):3–29.

Genton, M. and E. Ronchetti. 2003. "Robust Indirect Inference." *Journal of the American Statistical Association* 98(461).

Gourieroux, C. and A. Monfort. 1995. *Statistics and Econometric Models*. Cambridge University Press.

Gourieroux, C., A. Monfort and E. Renault. 1993. "Indirect Inference." *Journal of Applied Econometrics* 8:S85–S118.

Gourieroux, C., A. Monfort and E. Renault. 1996. "Two-Stage Generalized Moment Method with Applications to Regressions with Heteroscedasticity of Unkown Form." *Journal of Statistical Planning* 50:37–63.

Gourieroux, C., E. Renault and N. Touzi. 1999. *Calibration by Simulation for Small Sample Bias Correction.* in: Mariano, R., Schuermann, T., and Weeks, M., (Eds.), "Simulation Based Inference in Econometrics: Methods and Applications", Cambridge Univeristy Press.

Gross, S. and W. Steiger. 1979. "Least Absolute Deviation Estimates in Autoregressions with Infinite Variance." *Journal of Applied Probability* 16(2):104–116.

Hall, A. 2005. *Generalized Method of Moments.* Oxford University Press.

Han, Y. 2006. "Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model." *The Review of Financial Studies* 19:237–271.

Hannan, E. and M. Kanter. 1977. "Autoregressive Processes with Infinite Variance." *Journal of Applied Probability* 14(2):411–415.

Hansen, L.P. 1982. "Large Sample Properties of Generalized Methods of Moments Estimation." *Econometrica* 50:1029–1054.

Harvey, A.C., E. Ruiz and N. Shephard. 1994. "Multivariate stochastic variance models." *Review of Economic Studies* 61:247–264.

Hill, J. and E. Renault. 2008. "Kernel Self Normalized Tail Quantile Trimmed Sumds for Dependent, Heterogeneous Data, with Applications to Simulated GMM." Working Paper - UNC-CH.

Jacquier, E., N.G. Polson and P.E. Rossi. 1995. "Models and priors for multivariate stochastic volatility." Working Paper WP-95s-18 - CIRANO.

Kanter, M. and W. Steiger. 1974. "Regression and Autoregression with Infinite Variance." *Advances in Probability* 6(4):768–783.

Kim, S., N. Shephard and S. Chib. 1998. "Stochastic Volatility: likelihood inference and comparison with ARCH models." *Review of Economic Studies* 45:361–393.

King, M.A., E. Sentana and S.B. Wadhwani. 1994. "Volatility and Links Between National Stock Markets." *Econometrica* 62:901–933.

Kobayashi, M. and X. Shi. 2005. "Testing for EGARCH Against Stochastic Volatility Models." *Journal of Time Series Analysis* 26:135–150.

Lamoureux, C. and W. Lastrapes. 1990. "Persistence in Variance, Structural Change, and the GARCH Model." *Journal of Business and Economic Statistics* 8(2):225–234.

Lange, T., A. Rahbek and S. Jensen. 2007. "Estimation and Aysmptotic Inference in the First Order AR-ARCH Model." Working Paper - University of Copehagen.

Lanne, M. and P. Saikkonen. 2007. "A Multivariate Generalized Orthogonal Factor GARCH Model." *Journal of Business and Economic Statistics* 25:61–75.

Ling, S. 2007. "Self-Weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models." *Journal of Econometrics* 140:849–873.

Ling, S. and M. McAleer. 2003. "Asymptotic Theory for a Vector ARMA-GARCH Model." *Econometric Theory* 19:280–310.

Marquering, W. and M. Verbeek. 2004. "The Economic Value of Predicting Stock Index Returns and Volatility." *Journal of Finance and Quantitative Analysis* 39(2):407–429.

McFadden, D. and W. Newey. 1994. *Large Sample Estimation and Hypothesis Testing.* in: Engle, R., and McFadden, D. (Eds.), "Handbook of Econometrics 4", North-Holland.

McNeil, A., R. Frey and P. Embrechts. 2005. *Quantitative Risk Management.* Princeton University Press.

Meddahi, N. and E. Renault. 2004. "Temporal aggregation of volatility models." *Journal of Ecnometrics* 119:355–379.

Melino, A. and S.M. Turnbull. 1990. "Pricing foreign currency options with stochastic volatility." *Journal of Econometrics* 45:239–66.

Pakes, A. and D. Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57(5):1027–1057.

Pitt, M. and N. Shephard. 1999. *Time varying covariances: a factor stochastic volatility approach.* in: Bernardo, J.M., Berger, J.O. , David, A.P., and Smith, A.F.M., (Eds.), "Bayesian Statistics", Oxford University Press.

Ronchetti, E. and F. Trojani. 2001. "Robust Inference with GMM Estimators." *Journal of Econometrics* 101:37–69.

Rosenberg, B. 1972. *The Behavior of Random Variables with Nonstationary Variance and the Distribution of Security Prices.* in: Shephard, N., (Ed.), "Stochastic Volatility", Oxford University Press - 2005.

Shephard, N. 1996. *Statistical aspects of ARCH and stochastic volatility.* in: Cox, D.R., and Hinkley, D.V. , and Barndoff-Nielsen, O.E., (Eds.), "Time Series Models in Econometrics, Finance, and Other Fields", Chapman & Hall.

Spyropoulous, K., E. Kiountouzis and A. Young. 1973. "Discrete Approximation in the L^1 Norm." *Computer Journal* 16:180–186.

Stigler, S. 1973. "The Asymptotic Distribution of the Trimmed Mean." *The Annals of Statistics* 1(3):472–477.

Taylor, S. 1982. *Financial Returns Modeled by the Product of Two Stochastic Processes - A Study of the Daily Sugar Prices 1961-1975.* pp. 203–226. in: Andersen, O. (Ed.), "Time Series Analysis: Theory and Practice 1", Noth Holland.

Vetzal, K. 1992. Stochastic Short Rate Volatility and the Pricing of Bonds and Bond Options PhD thesis Faculty of Management, University of Toronto, Toronto, Canada.